

On the Role of Imitation in Social Networks

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Economics
at Universidad Carlos III de Madrid



Nikolaos Tsakas

Supervised by Antonio Cabrales

May 5, 2014

Acknowledgements

It was slightly more than five years ago when, sitting on my desk in Denmark, I was deciding to accept the offer from UC3M. Who on earth could imagine all these life-changing experiences that have covered every day until the moment I am writing these lines. So many special people, so many special memories, but, you know, if there is one thing that makes this adventure really unique, it is this rare moment of pure happiness where “it works”!!! It is all about this chill that covers you at the very moment that you understand the incomprehensible. There is only one problem; that usually, what you think you understood, it is wrong and what you think that worked, it does not. Those are the very moments that you really want to thank the people around you. Either for helping you understand what and why is wrong (although, honestly speaking, everyone hates for a second the guy who spots the mistake), or more importantly for turning your thoughts to different directions. For these and for many other reasons this is my only chance to thank all those that have been there during these last five year.

I would like to start this list by specially thanking my advisor Antonio Cabrales. For being always available and willing to listen patiently whatever I had to tell him, no matter how stupid or useless this was; For guiding me constantly out of dangerous dead-ends; For supporting me through the toughest moment of this era; But mainly because of his boundless passion for asking questions and trying to answer them, which I think that it has been the most important lesson I have ever been taught.

Moreover, I owe special thanks to Ignacio Ortuño for the countless conversations that we have shared over the past three years, each of which has been great food for thought. I should not forget to thank the rest of my teachers, who have tried patiently to make me learn economics, irrespectively of how much I have been resisting. Let us hope that the future will show that something has come out of all this.

A huge part of this letter should be dedicated to all my friends and colleagues during the last five years. As I have said in the past, it is always better when you feel that you are not alone on this long journey. I would need several pages to explain to each one of you separately how important your presence and friendship has been for me, so to keep it short I really thank you all from the bottom of my heart. The list is very long, but it should be unfair not to send my special gratitude to Christo, Pedro, Francesco, Marta, Albert, Fernando, Xiao, Emre, Mehdi, as well as Tasso and Pablo, Guillermo, Adrian and Kike. My best wishes to all of you guys and hope to be seeing all of you around.

One should never forget that how important are those that are so close no matter how far. Alexi C, Alexi K, Gianni, Kosta I feel blessed to have shared so many Sunday Skype afternoons and summer philosophical conversations with you.

But as they say, it always comes down to family. I owe special thanks to my parents Thoma and Roula for being as exaggeratedly supportive and protective as twenty five years ago and for teaching by example the satisfaction of selflessness. Even if I do not say it to you very often, I will always thank you for everything. Also, to Alexandra for being the person that keeps the balance in this family, since the moment that became (in practice) part of it. But most of all to my brother Elias, because I can hardly imagine how my life would have been if he hasn't been always there for me. Thanks φίλε.

And you always leave the sweet for the end...To Amalia, just for being my one and only most important person in this world.

Contents

1	Imitating the Most Successful Neighbor in Social Networks	8
1.1	Introduction	8
1.1.1	Motivation	8
1.1.2	Preview of Results	9
1.1.3	Related Literature	11
1.2	The Model	13
1.2.1	The agents	13
1.2.2	The Network	14
1.2.3	Behavior	14
1.2.4	Steady state and efficiency	15
1.3	Networks with finite population	15
1.3.1	Experimentation	17
1.3.2	Probabilistic updating	20
1.4	Networks with arbitrarily large population	20
1.4.1	Bounded Neighborhoods	21
1.4.2	Unbounded Neighborhoods	24
1.5	Conclusion	26
1.6	Appendix	27
2	Diffusion by Imitation: The Importance of Targeting Agents	35
2.1	Introduction	35
2.1.1	Motivation	35
2.1.2	Setting and Results	37
2.1.3	Related Literature	39
2.2	The Model	41
2.2.1	The Agents	41
2.2.2	The Network	42
2.2.3	The Behavior	42
2.2.4	The Problem	43
2.3	Results for an Impatient Planner	44

2.4	Results for an Infinitely Patient Planner	45
2.4.1	Preliminaries	46
2.4.2	Main Results	47
2.5	Extensions	51
2.6	Conclusion	53
2.7	Appendix A - Extensions	54
2.7.1	Intermediate patience - the τ -patient planner	54
2.7.2	Endogenous number of targeted agents	57
2.7.3	Expected Waiting Time before Absorption	59
2.7.4	$q \neq 1 - p$: Conformity, Switching Cost and Possibility of Inertia	64
2.7.5	On the Importance of Centrality – The Line	66
2.7.6	On the Importance of Centrality – The Star	69
2.7.7	On the Importance of Centrality – Other Networks	70
2.8	Appendix B - Proofs	72
2.9	Appendix C - Figures	93
2.10	Appendix D - Matlab Code	97
3	Optimal Influence under Observational Learning	102
3.1	Introduction	102
3.1.1	Motivation	102
3.1.2	Setting	103
3.1.3	Results	104
3.2	Model	105
3.2.1	The Network	105
3.2.2	The Problem	106
3.3	Results: Targeting only one agent	107
3.4	Results: Targeting several agents	109
3.4.1	The Circle	112
3.5	Conclusion	118
3.6	Appendix	118

Introduction

This thesis focuses on the role of imitation in social networks. We study diffusion processes where the agents imitate successful past behavior of their neighbors. I characterize the long run properties of these processes, as well as targeting strategies that lead to maximum diffusion of a specific desired action in a network. Although theoretical, the results can have direct applications in different fields, like marketing and public policy.

Chapter 1. We study a model of observational learning in a set of agents who are connected through a social network. The agents face identical decision problems under uncertainty and update their choices myopically, imitating the choice of their most successful neighbor. We show that in finite networks, regardless of the network structure, the population converges to a monomorphic steady state, i.e. one at which every agent chooses the same action, and it cannot be predicted which this action will be. In arbitrarily large networks with bounded neighborhoods, an action is diffused to the whole population either if it is the only one initially chosen by a non-negligible share of the population, or if the payoffs satisfy a sufficient condition. Without the assumption of bounded neighborhoods, (i) an action can survive even if only one agent chooses it initially, and (ii) there may exist steady states that are not monomorphic.

Chapter 2. We study the optimal targeting strategy of a planner who seeks to maximize the diffusion of an action in a circular network where agents imitate successful past behavior of their neighbors. We find that the optimal targeting strategy depends on two parameters: (i) the likelihood of the action being more successful than its alternative and (ii) the planner's patience. More specifically, when the planner's preferred action has higher probability of being more successful than its alternative, then the optimal strategy for an infinitely patient planner is to concentrate all the targeted agents in one connected group; whereas when this probability is lower it is optimal to spread them uniformly around the network. Interestingly, for a very impatient planner, the optimal targeting strategy is exactly the opposite. Our results highlight the importance of knowing a society's exact network structure for the efficient design of targeting strategies, especially in settings where the agents are positionally similar.

Chapter 3. We study a problem of optimal influence in a society where agents learn from their neighbors. We consider a firm that seeks to maximize the diffusion of a new product whose quality is ex-ante uncertain, to a market where consumers are able to compare the qualities of two alternative products as soon as they observe both of them. The firm can seed the product to a subset of the population and our goal is to find which is the optimal subset to be targeted. We provide a necessary and sufficient condition that fully characterizes the optimal targeting strategy for any network structure. The key parameter in this condition is the agents' decay centrality, which is a measure that takes into account how close an agent is to others, but in a way that very distant agents are weighted less than closer ones.

Chapter 1 is dedicated to Amalia

Chapter 1

Imitating the Most Successful Neighbor in Social Networks

1.1 Introduction

1.1.1 Motivation

Social learning theory is a perspective that states that social behavior is learned primarily by observing (observational learning) and imitating the actions of others (see Ormrod, 1999). In our context, social learning describes the idea that economic agents' decisions are influenced by past experience of their neighbors.¹ We consider a problem of individual decision-making under uncertainty, where the agents are not aware of the relative profitability of their alternative choices. In this setting, we introduce a simple learning process with the following features: Agents observe in each period the choices of their neighbors and the payoffs those choices yield. Subsequently, they revise their choices repeatedly according to these observations. In particular, they do so by imitating the action that yielded the highest payoff within their neighborhood in the preceding period. We refer to this updating rule as "imitate-the-best" neighbor.² The aim of this paper is to analyze the long-run behavior of a population consisting of agents who behave as if they "imitate-their-best" neighbor.

Imitation of successful behavior has been studied extensively in several different disciplines.³ In economics, recent empirical results suggest that people tend to imitate successful past behavior of their neighbors (see Apesteguia et al., 2007; Conley and Udry, 2010; Bigoni and Fort, 2013). All these articles provide evidence that in several dynamic decision problems, the agents seem to behave as if they observe the actions of their neighbors, and they tend to imitate those who were the most successful in the recent past. There are several reasons that can justify this behavior. On the one hand, agents may not be aware of the mechanisms controlling the outcome of their choices, hence they need to experiment themselves or rely on past experience of those they can observe. On the

¹A similar definition is used in (Ellison and Fudenberg, 1993).

²We have borrowed the term from Alós-Ferrer and Weidenholzer (2008).

³In psychology see Bandura (1977) and Ormrod (1999) and in animal behavior see Zentall (2003).

other hand, in certain environments, Bayesian updating may require calculations that are beyond the computational capabilities of the agents, leading them to adopt simple learning heuristics.

There are several real environments that can be described by our setting. An example where our model fits particularly well is the diffusion of agricultural technologies.⁴ It is reasonable to assume that farmers are able to observe the technologies used by their neighbors, as well as the output those technologies yield. Moreover, Conley and Udry (2010) provide evidence supporting the idea that farmers tend to adopt technologies which have performed extremely well in near past. Furthermore, the output of a production activity is subject to both aggregate and idiosyncratic stochastic shocks. On the one hand, aggregate shocks are related to the appropriateness of the weather for different crops and technologies. On the other hand, idiosyncratic shocks are related to the specific characteristics of the terrain of each farm, the efficient, or not, implementation of a given technology and several other factors of unobserved heterogeneity between farmers and their land. Notice that, our focus is on the output produced by a given technology and not on the profits this output generated. This fact allows us to disregard completely issues related to the market structure and competition among farmers. Therefore, it becomes reasonable to use a model without strategic interactions between farmers.

Another relevant example is the parents' decision about which school to send their children to; or their decision about whether to send them to a public school or a private one. It is apparent that the satisfaction of the parents by such a decision depends mainly on the characteristics of the school itself, rather than on the decisions of other parents. Nevertheless, it may be different between parents who made the same choice. It is also commonly observed that the parents decide relying mostly on other parents' previous experience. This happens mainly because of the difficulty lying on the identification of the real quality of each school. Furthermore, information received by those who had been extremely satisfied in the past tends to be more influential; observation that leads directly to our learning heuristic.

1.1.2 Preview of Results

Formally, we consider a countable population forming a social network. In each period, every agent chooses an action from a finite set of alternatives. Payoffs are uncertain and depend on the action chosen and on two uncertainty parameters. The first parameter is associated with an idiosyncratic shock that is independent across agents. The second parameter is associated with an aggregate shock that is common for all the agents who have chosen the same action and is independent across actions. All shocks are independent across periods and every action has strictly positive probability of yielding higher outcome than all the other actions. The agents are not aware of the underlying distributions of the uncertainty parameters and there are no strategic interactions between agents. In the example of diffusion of agricultural technologies this payoff structure is very intuitive. Output levels depend partially on aggregate stochastic shocks, mostly due to the weather, but they depend

⁴For further references see also Ellison and Fudenberg (1993).

also on idiosyncratic shocks that capture differences in the soil and climate or even in the proper use of a technique. After making their own choice, all agents observe the chosen actions and realized payoffs of their direct neighbors. Subsequently, they update their choice, imitating myopically the action that yielded the highest payoff within her neighborhood in the preceding period.

We show that when the population is finite the network eventually converges with probability one to a monomorphic steady state, meaning that all the agents end up choosing the same action. However, we cannot guarantee which action will be the one to survive. This happens because each action is vulnerable to a sequence of negative shocks that can lead to its disappearance. We extend our analysis, introducing a standard form of experimentation, where with some probability agents choose randomly among the alternative actions, instead of imitating their most successful neighbor. Interestingly enough, despite the fact that experimentation allows the reappearance of actions that have disappeared, if the experimentation rate becomes arbitrarily small the process stays almost always in one of the monomorphic states. Moreover, for each one of the monomorphic states the probability of the process being in this state is strictly positive. This means that in the long run the process will spend a non-negligible amount of time in each one of the monomorphic states. Thus, we can say that the result is also robust to the standard form of experimentation.

Our result is shaped mainly by a combination of the two main ingredients of the model: the imitative behavior of the agents and the payoff structure. The finite population combined with the fact that each action may yield higher or lower payoff than its alternatives make any action vulnerable to disappearance after a sufficiently long sequence of negative shocks. This, in turn, means that is possible the contagion of a sub-optimal action. The possibility of convergence to a state where everyone chooses a sub-optimal action is a common issue in processes with stochastic payoffs (see for instance Bala and Goyal, 1998; Gale and Kariv, 2003). However, the crucial difference is that in our process it is impossible to eliminate this possibility, due to the imitative behavior of the agents. The fact that the agents do not accumulate information regarding the underlying distributions of payoffs makes impossible the prediction of their long-run behavior.

Notice that this is impossible even under the presence of experimentation. This is in contrast to several of the models in which the agents interact strategically with their neighbors and where experimentation can ensure convergence to the efficient action (see for instance Ellison, 1993; Alós-Ferrer and Weidenholzer, 2008).⁵ The fact that a single mutation can generate a transition from one monomorphic state to another ensures that the process will spend a non-trivial amount of time at each one of its monomorphic states. In principle, for finite populations these results can be extended to other imitation rules, such as "imitate best average" (see Ellison and Fudenberg, 1993).⁶

The results differ significantly when we let the population become arbitrarily large. First of all,

⁵In our setting we should be careful when referring to efficiency. Referring to an action by using the term efficient might be misleading, since we have not defined precisely the payoffs' distributions. For this reason, throughout the paper we use the term optimal, which will refer to an action being more probable to yield higher payoff than the alternative actions.

⁶Where an agent chooses the action she observes to have yielded the highest average payoff in her neighborhood during the preceding period.

without further restrictions we cannot ensure convergence to a monomorphic steady state. In fact, convergence depends on the payoffs' distributions, on the network structure and in particular on whether or not the agents have bounded neighborhoods, i.e. whether there are agents who interact with a non-negligible share of the population.⁷ The importance of neighborhoods being bounded arises in several models of local interactions and usually the existence of agents who can affect a large proportion of the population has a negative effect on long-run behavior of the population.⁸

Assuming bounded neighborhoods, we can ensure convergence to a monomorphic state if there is only one action chosen initially by a non-negligible share of the population. If this is not the case, then we provide a counter example, where a network never converges to a steady state. Nevertheless, we provide a sufficient condition on the payoff structure, which ensures convergence regardless of the network structure. This condition is more demanding than first order stochastic dominance, thus implying that in very large networks the diffusion of a single action is very hard to occur and it demands a very large proportion of initial adopters, or a special network structure, or an action to perform much better compared to all others. This is the only result where we are able to ensure convergence to a specific action. Our sufficient condition disregards completely the importance of the network architecture. The behavior of specific network structures would be a very interesting topic for further research.

Once we drop the assumption of neighborhoods being bounded, the properties of the network change significantly. In this case, an action may survive, even if only one agent chooses it initially. This happens because this one agent may affect the choice of a non-negligible share of the population; an observation that stresses the role of centrality in social networks. For instance, providing a technology or a product to a massively observed agent, can affect significantly the behavior of the population. Finally, we construct another example where a network is in steady state without this being monomorphic, which contrasts our result regarding networks with finite population.⁹

1.1.3 Related Literature

Our work is in line with the literature on learning from neighbors (see Banerjee, 1992; Ellison and Fudenberg, 1993, 1995; Bala and Goyal, 1998; Gale and Kariv, 2003; Banerjee and Fudenberg, 2004; Acemoglu et al., 2011). Most of this literature has focused on the identification of conditions that ensure the contagion of efficient actions to the whole population, typically in settings without strategic interactions. A similar environment to ours is considered by Ellison and Fudenberg (1993), where the agents choose repeatedly between two technologies and evaluate their choices periodically. The authors consider an "imitate-the-best-average" rule, rather than "imitate-the-best", and they focus

⁷By bounded neighborhood we mean that there exists $K > 0$ such that the number of neighbors of every agent i satisfies $k_i \equiv |N_i| \leq K, \forall i \in N$.

⁸See for instance Bala and Goyal (1998) and Golub and Jackson (2010). Despite the fact that the settings are substantially different, the effect of such agents is evident.

⁹Coexistence of several actions in the long-run has also been observed in other papers, as for instance in Ellison and Fudenberg (1993).

exclusively on infinitely large populations, thus restricting the comparability of our results.

Of particular interest is the paper by Gale and Kariv (2003), where the authors consider the same network and payoff structure as we do, but they assume that agents perform Bayesian updating on their potentially different initial beliefs.¹⁰ Updating is based on the observed choices and realized payoffs of their neighbors. Similarly to our results, they find that asymptotically all agents will converge to the adoption of the same action. However, they are able to ensure convergence to the optimal action under certain conditions, which in our setting is never possible for networks with finite population.

Nevertheless, the main focus of our paper is the feature of imitation of successful behavior. Large part of the literature on imitation in networks studies games played between neighbors and usually either versions of prisoner's dilemma or coordination games (see for instance Alós-Ferrer and Weidenholzer, 2008; Ellison, 1993; Eshel et al., 1998). In particular, imitation of the most successful neighbor is studied by Alós-Ferrer and Weidenholzer (2008) in the context of a coordination game played by agents located on an arbitrary network. The main aim of these papers is either how to sustain cooperation or how to achieve coordination to the most efficient action. As we have already mentioned, this is something that can never be guaranteed in our case. This happens mainly due to the fact that in these models uncertainty arises from the lack of information about the choices of others. Disregarding that, the transition between action profiles occurs deterministically. This feature allows the characterization of conditions that ensure contagion to efficient action or sustain cooperation, using for instance results such as the ones in Ellison (2000).

Imitative behavior is also discussed in several models of evolutionary game theory (see Weibull, 1995; Fudenberg and Levine, 1998), where in fact imitation tends to be a particularly efficient form of behavior. More specifically, Vega-Redondo (1997) shows that if a Cournot economy consists of agents who imitate the most successful agent they observe, then it is led to Walrasian equilibrium. Furthermore, Schlag (1998, 1999) finds probabilistic imitation of successful agents to be the most efficient behavior when facing multi-armed bandits. The main difference between evolutionary and network models, is that agents are randomly matched with others from the population.

Finally, imitation has been studied extensively in several different disciplines, including Physics, Computer Science, Biology and Zoology. In particular, Nowak and May (1992) may be considered as one of the first papers developing the idea of imitating the most successful neighbor. Many of the papers in this literature (see for instance Nowak and May, 1993; Abramson and Kuperman, 2001; Nowak et.al, 2004) study games played between neighbors and they focus mostly on the evolution of cooperation.¹¹ Environments without strategic interactions are considered in voter models (see Liggett, 1985) and in general models of cellular automata. The difference is that, in the latter settings, imitation is not associated with successful past behavior of one's neighbor, but it rather

¹⁰Not surprisingly, a large part of the literature on learning from neighbors has focused on Bayesian learning and best-response strategies, rather than on imitation (see for example Bala and Goyal, 1998; Gale and Kariv, 2003; Acemoglu et al., 2011).

¹¹For an extensive survey on evolutionary games on graphs see Szabó and Fáth (2007).

describes a random choice among observed actions.

The rest of the paper is organized as follows. In Section 1.2, we explain the model. In Section 1.3, we provide the main results for networks with finite population. While, in Section 1.4 we study networks with arbitrarily large population. Finally, in Section 1.5 we conclude and discuss possible extensions.

1.2 The Model

1.2.1 The agents

There is a countable set of agents N , with cardinality n and typical elements i and j , mentioned as *population* of the network.¹² Each agent $i \in N$ takes an action a_i^t from a finite set $A = \{\alpha_1, \dots, \alpha_M\}$, at every period $t = 1, 2, \dots$. Each $a \in A$ yields a random payoff.

Uncertainty is represented by a probability space $(S, \mathcal{F}, \mathbb{P})$, where $S = \Omega \times Z$ is a product metric space, with Ω and Z being compact metric spaces¹³. \mathcal{F} is the product Borel σ -algebra, and \mathbb{P} is a product Borel probability measure.

There is a common stage payoff function $U : A \times \Omega \times Z \rightarrow \mathbb{R}$, which consists of two components, U_{ID} and U_{AC} . The first component, U_{ID} , is associated with an idiosyncratic shock, $\omega \in \Omega$, which is realized every period independently for each agent and its distribution depends on the chosen action. The second component, U_{AC} , is associated with an aggregate shock, $z \in Z$, which in each period is common for all the agents who chose the same action in that period. Its realization is independent across actions and across periods. Given the meaning of the payoffs in the current setting, we can assume without problems the payoff function to be additively separable in these two components. Our results would be identical without this assumption and we impose it only because it facilitates the exposition of our results.

The payoff structure is similar to the ones used by Gale and Kariv (2003) and Ellison and Fudenberg (1993). A simpler form of payoffs would lead to similar results, nevertheless with the current setting we are able to capture several realistic environments, such as the outcome of the adoption of different agricultural technologies. Formally,

$$U(a, \omega, z) = U_{ID}(a, \omega) + U_{AC}(a, z)$$

where both $U_{ID} : A \times \Omega \rightarrow S_{ID} \subset \mathbb{R}$ and $U_{AC} : A \times Z \rightarrow S_{AC} \subset \mathbb{R}$ are bounded and continuous in Ω and Z respectively. $U_{ID}(a, \Omega) := \{x \in \mathbb{R} : \exists \omega \in \Omega \text{ s.t. } U(a, \omega) = x\}$ and $U_{AC}(a, Z) := \{x \in \mathbb{R} : \exists z \in Z \text{ s.t. } U(a, z) = x\}$ are closed intervals in \mathbb{R} .¹⁴ We restrict our attention to cases where $U_{ID}(a, \Omega) = U_{ID}(a', \Omega)$ and $U_{AC}(a, Z) = U_{AC}(a', Z)$ for all $a, a' \in A$. Moreover, \mathbb{P} has full support over $\Omega \times Z$. We denote by F_{ID}^a and F_{AC}^a the cumulative distribution functions of U_{ID} and U_{AC} ,

¹²In Section 3 we assume that n is finite, whereas in Section 4 we assume n to be arbitrarily large.

¹³The product of two compact metric spaces is also a compact metric space.

¹⁴Throughout the paper, $U_{ID}(a, \Omega)$ is called the (common) support of the idiosyncratic component of the payoff function and likewise $U_{AC}(a, Z)$ is called the (common) support of the aggregate component of the payoff function.

respectively, associated with action a .¹⁵ This setting ensures that for each action there is positive probability to be the one to yield the highest outcome in each period.¹⁶

In each period t , the realized payoff of agent $i \in N$ who has chosen action $a_i^t \in A$ is denoted by U_i^t . More specifically, for ω_i^t being the realization of the idiosyncratic shock of player i and $z_{a_i^t}^t$ being the shock associated to the action chosen by agent i , both at period t , we obtain the following expression:

$$U_i^t = U_{ID}(a_i^t, \omega_i^t) + U_{AC}(a_i^t, z_{a_i^t}^t)$$

1.2.2 The Network

A social network is represented by a family of sets $\mathcal{N} := \{N_i \subseteq N \mid i = 1, \dots, n\}$, with N_i denoting the set of agents observed by agent i . Throughout the paper N_i is called i 's neighborhood, and is assumed to contain i .¹⁷ The sets \mathcal{N} induce a graph G with nodes N , and edges $E = \bigcup_{i=1}^n \{(i, j) : j \in N_i\}$. We focus on undirected graphs: as usual, we say that a network is undirected whenever for all $i, j \in N$ it is the case that $j \in N_i$ if and only if $i \in N_j$. In the present setting, the network structure describes the channels of communication in the population and does not impose strategic interactions. More specifically, each agent $i \in N$ observes the action and the realized payoff of every $j \in N_i$.

A path in a network between nodes i and j is a sequence i_1, \dots, i_K such that $i_1 = i$, $i_K = j$ and $i_{k+1} \in N_{i_k}$ for $k = 1, \dots, K - 1$. The distance, l_{ij} , between two nodes in the network is the length of the shortest path between them. The diameter of the network, denoted as d_N , is the largest distance between any two nodes in the network. We say that two nodes are connected if there is a path between them. The network is connected if every pair of nodes is connected. We focus on connected networks, nevertheless for disconnected networks the analysis would be identical for each of their connected components.¹⁸

1.2.3 Behavior

In the initial period, $t = 1$, each agent is assigned, exogenously, to choose one of the available actions.¹⁹ After each period, $t = 2, 3, \dots$ the agents have the opportunity to revise their choices. Revisions happen simultaneously for all agents. We assume that each agent $i \in N$ can observe the choices and the realized payoffs of her neighbors in the previous period. According to these

¹⁵Without the assumption of U being additively separable we would have to do exactly the same analysis but using the joint cumulative distribution functions.

¹⁶The cases where an action is strictly payoff dominant are uninteresting, because it is trivial that the dominant action will spread to the whole population. In Section 3 we explain how this result would be affected by experimentation.

¹⁷This assumption is not usual, however in this setting it is necessary since N_i describes the set of agents whose actions can be observed by i . Therefore, since it is reasonable to assume that one can observe her own actions, she should be contained in her own neighborhood.

¹⁸A component is a non-empty sub-network N' such that $N' \subset N$, N' is connected and if $i \in N'$ and $(i, j) \in E$ then $j \in N'$ and $(i, j) \in E'$.

¹⁹We assume that every action in A is chosen initially by some agent. Without loss of generality we can exclude from the action space any action that is not chosen by any agent during the first period.

observations, she revises her choice by imitating the action of her neighbor who received the highest payoff in the previous period. Ties are broken randomly.²⁰ Formally, for $t > 1$,

$$a_i^{t+1} = a_k^t \text{ with } k \in \arg \max_{j \in N_i} U_j^t \quad (1.1)$$

where recall that by a_k^t we denote the action chosen by agent k in period t and by U_j^t the realized payoff of agent j in period t . We refer to this updating rule as “imitate-the-best” (see also Alós-Ferrer and Weidenholzer, 2008).

An important aspect of this myopic behavior is that the agents discard most of the available information. They ignore whatever happened before the previous round and even from this round they take into account only the piece of information related to the most successful agent. This naive behavior makes the network vulnerable to extreme shocks, which may be very misleading for the society.

1.2.4 Steady state and efficiency

State of period t is called the vector (a_i^t, \dots, a_n^t) of the actions chosen by each agent at this period, which belongs to the state space A^n . We denote by $A_t = \{\alpha_k \in A : \exists i \in N \text{ such that } a_i^t = \alpha_k\}$ the subset of the action space, A , which contains those actions that are chosen by at least one agent in period t . Notice, that an action which disappears from the population at a given period, never reappears, hence $A_t \subseteq A_{t-1} \subseteq \dots \subseteq A_1 \subseteq A$. In a given period t , its state is called *monomorphic* if every agent chooses the same action, i.e. if there exists $k \in \{1, \dots, M\}$, such that $a_i^t = \alpha_k$ for all $i \in N$. Also, the population is in *steady state* if no agent changes her action from this period on, i.e. if $(a_i^t, \dots, a_n^t) = (a_i^{t'}, \dots, a_n^{t'})$ for all $t' > t$. Throughout the paper, the idea of convergence refers to *convergence with probability one*. Finally, we call an action *optimal* if it is the most likely to yield the highest payoff compared to all the other actions. An action is *better* than another action if it more likely to yield higher payoff. An action is called *sub-optimal* if it is not better than all the other.

1.3 Networks with finite population

In this section, we restrict our attention to networks with finite population. We prove that finite networks always converge to a monomorphic steady state, regardless of the initial conditions and the network structure. Moreover, any action can be the one to survive in the long run.

Before presenting our main results, it is worth mentioning a remark regarding the complete network. In a complete network, each agent is able to observe actions and realized payoffs of every other agent, i.e. $N_i = N$ for all $i \in N$. If the network is complete, then it will converge to a monomorphic steady state from the second period on. Namely, everyone else in the second period will imitate the agent who received the highest payoff in the first period, leading to the disappearance

²⁰Ties arise with probability zero, because the agents are countably many, while the states of nature are uncountably many.

of all alternative actions. The probability of two actions giving exactly the same payoff is equal to zero, because of the assumptions regarding the continuity of payoff functions' distributions.

Obviously, the simple case mentioned above does not ensure neither the convergence to a steady state, nor that this steady state needs to be monomorphic. The following proposition establishes the fact that if the system reaches a steady state, then it must be the case that all the agents choose identically. Formally,

Proposition 1. *Consider an arbitrary connected network with finite population. If the agents behave under imitate-the-best rule, then all possible steady states are necessarily monomorphic.*

The proof (which can be found in the appendix) is very intuitive. The idea is that when more than one actions are still chosen, there must exist at least one pair of neighbors choosing different actions. Hence, each one of them faces a strictly positive probability of choosing a different action in the next period. This ensures that at some period in the future, at least one of the two will change her action, meaning that the population is not in steady state. The above proposition is in line with the work of Gale and Kariv (2003), where identical agents end up making identical choices in the long run. Notice that the result is no longer valid if the network becomes arbitrarily large (see Example 5 - Two Stars). Nevertheless, we have not ensured yet the convergence of the population to a steady state, but only the fact that if there exists a steady state, then it has to be monomorphic. In principle, all monomorphic states are possible steady states.

In the following theorem we provide the first main result of the paper, which is that any arbitrary network, with finitely many agents, who behave under "imitate-the-best rule", converges to a monomorphic steady state. The intuition behind our result is captured by the following three lemmas, which are also used for the formal proof (all proofs can be found in the appendix).

Lemma 1. *Consider an arbitrary connected network with finite population, where more than one action is observed. Each period t , every action $\alpha_k \in A_t$ faces positive probability of disappearing after no more than d_N periods.*

The main idea behind this lemma is that every action is vulnerable to a sequence of negative shocks to the payoffs of its adopters (agents who are currently choosing the action). Regardless of the number and the position of those agents, the probability of disappearance after finitely many periods is strictly positive.

Lemma 2. *Suppose that $K - 1$ actions have disappeared from the network until period t . Then, there is strictly positive probability that exactly K actions will have disappeared from the network after a finite number of periods τ .*

This result is a direct implication of Lemma 1. Its importance becomes apparent if we notice that an action that disappears from the population (not chosen by any agent at a given period) never reappears. τ can take different values depending on the structure of the network and the initial conditions, but it is always bounded above by the diameter of the network d_N .

Lemma 3. *Suppose that $K - 1$ actions have disappeared from the network until period t . Then there is strictly positive probability of convergence to a monomorphic state in the next $T = (M - K + 1)d_N$ periods.*

There are many possible histories that lead to convergence to a monomorphic state. Some of them lead to the disappearance of one action every $\tau = d_N$ periods, which is an intersection of events with strictly positive probability of occurring, as shown in Lemma 2.

Theorem 1. *Consider an arbitrary connected network with finite population. If the agents behave under "imitate-the-best" rule, then the network will converge with probability one to a monomorphic steady state.*

By Lemma 3 we know that convergence will occur with strictly positive probability after a finite number of periods. Analogously, convergence will not occur in the same number of periods with probability bounded below one. The probability that convergence never occurs is given by the infinite product of probabilities strictly lower than one, therefore convergence to a monomorphic steady state is guaranteed in the long run.

Corollary 1. *Consider an arbitrary connected network with finite population. If the agents behave under "imitate-the-best" rule, then there is always strictly positive probability of convergence to a sub-optimal action.*

The corollary is apparent from the fact that even the optimal action faces a positive probability of disappearance as long as there are more actions chosen in the network. This result points out a weakness of this updating mechanism, which is the inability to ensure efficiency. However, if the population is arbitrarily large, then we provide sufficient conditions for the diffusion of the optimal action (see Section regarding Networks with Arbitrarily Large Population).

1.3.1 Experimentation

Although we have shown that all monomorphic states are steady states, this does mean that they are all equally plausible. It is important to discuss whether they are robust to small mutations and experimentations. Recall that one of the crucial features of our mechanism is that once an action disappears from the network it never reappears. Obviously, if we introduce forms of experimentation that preserve this feature, then the result will still hold. For example, we could transform the updating rule as follows: In each period, each agent imitates her most successful neighbor with probability $(1 - \epsilon) \in (0, 1)$ and with probability ϵ imitates at random another of the actions she observed in the previous period (including her own current choice). Under this revised updating rule, the result remains the same, with a slight modification in the lower bound of Lemma 1 which should be multiplied by $(1 - \epsilon)^{|I_k^t|}$.

However, one might find this definition of experimentation rather unusual. For this reason, we focus on the standard form of experimentation where in each period, each agent imitates her most

successful neighbor with probability $1 - \epsilon$ and with probability ϵ chooses at random among all the available actions. Applying the results of the seminal article by Fudenberg and Imhof (2006) we obtain the following two results. First, we show that the process will be almost always nearby one of the monomorphic states. Notice that, under this form of experimentation actions that have disappeared can reappear in the population. This means that there is no convergence to a steady state in the sense that we have defined it earlier, i.e. there is no state such that if the process reaches it, then stays there forever. Second, we describe an algorithm that calculates how much time the process is expected to spend in each of the monomorphic states. In fact, we observe that the process spends a non-negligible amount of time in each one of them.²¹

Fudenberg and Imhof (2006) observe that as long as the probability that a single mutant invades the population is $O(\epsilon)$, while the probability that two or more mutants invade simultaneously is $o(\epsilon)$, then the process will spend $1 - O(\epsilon)$ of the time at the steady states²² of the unperturbed process, which are the vertices of the state space, $O(\epsilon)$ of the time on the edges of the state space and $o(\epsilon)$ at the interior points. In our case, the set of steady states coincides with the set of monomorphic states and it is easy to see that the present form of experimentation verifies the above stated condition. Therefore, if the rate of experimentation is small then the process will be almost always nearby one of the monomorphic states.

Moreover, the current process verifies the assumptions of Theorem 2 of Fudenberg and Imhof (2006). This theorem provides an algorithm that calculates how much time the perturbed process is going to spend in each one of the steady states of the unperturbed process. The three conditions that are sufficient to verify the assumptions of the theorem are the following: (i) The fact that our form of experimentation satisfies the conditions defined in the previous paragraph,²³ (ii) our Theorem 1, in which we show that the set of steady states of the unperturbed process coincides with the set of monomorphic states and (iii) the fact that a single mutation can cause a transition from any steady state to any other.²⁴ Intuitively, the result states that it is sufficient to calculate the invariant distribution of a Markov matrix over the monomorphic (absorbing, in general) states where the probability of a transit from the state where “all choose A ” to the state where “all choose B ” is the transition probability from the state where “all but one choose A and one chooses B ” to the state where “all choose B ”.²⁵ It is important to notice that the process will spend a non-negligible amount of time in each one of the monomorphic states. The following example illustrates the result we just described.

²¹Usually, such states are called stochastically stable (see Ellison, 2000). Therefore, using this terminology, we could say that in the main theorem we have proven that the set of monomorphic states coincides with the set of absorbing states and in this section we show that all monomorphic states are also stochastically stable.

²²Mentioned by the authors as absorbing states.

²³For expositional reasons we use Assumptions 3 and 4 of Fudenberg and Imhof (2006), which are more intuitive and which combined imply Assumption 8 that is used in the proof of Theorem 2.

²⁴As an attempt to keep notation simple, we do not define the arguments and assumptions of the theorem formally. We refer the reader to the original article for the formal analysis.

²⁵Whenever possible we adopt the exact terminology of the original paper, in order to facilitate the reader.

Example 1. Consider the following circle network of n agents (Figure 1.1 - Circle Network) and assume that the action space consists of two actions α_1 and α_2 .

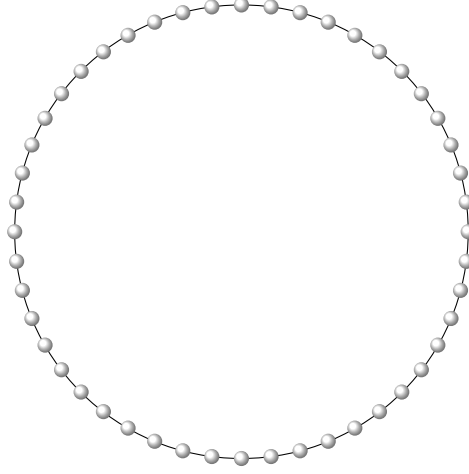


Figure 1.1: Circle Network with n agents.

For simplicity, assume that the individual shock for each agent is always equal to zero, i.e. $U_{ID}(a, \omega) = 0$ for all $a \in A$ and $\omega \in \Omega$. Therefore, the payoff is determined completely by the function U_{AC} . If in each period the probability that action α_1 yields higher payoff than action α_2 is $p \neq \frac{1}{2}$, then we can calculate the probability $P_{2 \rightarrow 1}$ ($P_{1 \rightarrow 2}$ respectively) that a population of agents choosing α_2 (α_1) invaded by a single agent choosing α_1 (α_2), as well as the relative time T_1 and T_2 that the process will spend in each one of the two monomorphic states. In particular, using standard results of finite Markov chains (see Kemeny and Snell, 1960, p. 153), for n being odd number and defining $r = \frac{p}{1-p}$, we get:

$$P_{2 \rightarrow 1} = \frac{r^{\frac{n-1}{2}-1}(r-1)}{r^{\frac{n-1}{2}}-1}, \quad P_{1 \rightarrow 2} = \frac{r-1}{r^{\frac{n-1}{2}}-1} \Rightarrow T_1 = \frac{P_{2 \rightarrow 1}}{P_{1 \rightarrow 2} + P_{2 \rightarrow 1}} = \frac{r^{\frac{n-1}{2}-1}}{r^{\frac{n-1}{2}}-1+1}, \quad T_2 = \frac{P_{1 \rightarrow 2}}{P_{1 \rightarrow 2} + P_{2 \rightarrow 1}} = \frac{1}{r^{\frac{n-1}{2}-1}+1}$$

It is important to mention what would happen under the presence of experimentation if instead of assuming random and independently distributed payoffs we had assumed the payoffs to be deterministic. First of all, it is obvious that without experimentation (and assuming the initial state to contain all the actions of the action space) the process would converge to the monomorphic state where all the agents choose the optimal action. The reason is that the optimal action would always yield higher payoff than any other action, therefore every agent would switch to the optimal action after observing it for the first time and would stick to it from that period onwards. The result would not change if we add experimentation, but let the experimentation rate go to zero. Proving this would be a direct application of Theorem 1 of Ellison (2000) (Simple Radius – Coradius Theorem). This becomes clear if we observe that in the unperturbed process all the non-monomorphic states where at least one agent chooses the optimal action lie in the basin of attraction of the monomorphic state where everyone chooses the optimal action. However, if we assume the experimentation rate to

be bounded away from zero, the process will enter a chaotic phase where all actions will be present but most of the agents will be choosing the optimal action. Intuitively, the dynamics of the process will be the following. On the one hand, there will be a natural tendency of the process towards the monomorphic state where everyone chooses the optimal action. On the other hand, the more agents choose the optimal action, the larger the number of them that will experiment by choosing an alternative action. The exact proportions would be hard to determine because they depend both on the experimentation rate and on the exact network structure.

Concluding, despite the fact that experimentation allows the reappearance of actions that have disappeared, we have shown that if the experimentation rate is low then the process will be almost always nearby one of the monomorphic states. Moreover, interestingly enough, the amount of time that the process will spend in each one of the monomorphic states is going to be strictly positive. Therefore, although we expect the process to spend more time at the monomorphic state where everyone chooses the optimal action, it need not be the case that this will be observed infinitely more often than the rest of the monomorphic states.

1.3.2 Probabilistic updating

Our results still hold if we relax the assumption that all the agents update their choices in every period and introduce probabilistic updating. Formally, suppose that in every period there is a positive probability, $r > 0$, that an agent will decide to update her choice. The proof is identical to the one of Theorem 1, if we multiply the lower bound of Lemma 1 by $r^{|I_k^t|}$, where $|I_k^t|$ is the cardinality of the set of agents choosing action α_k in period t . Hence, the network converges with probability one to a monomorphic steady state, although convergence may occur at a slower rate.

1.4 Networks with arbitrarily large population

In this section, we consider an arbitrarily large population N . Formally, what we mean is a sequence of networks $\{G_n\}_{n=1}^\infty = \{(\{1, \dots, n\}, \mathcal{N}^n)\}_{n=1}^\infty$, where every agent i of the n -th network is also an agent of the $(n+1)$ -th network, and moreover any pair of agents i, j of the n -th network are connected if and only if they are connected in the $(n+1)$ -th network. Roughly speaking, that would mean that the $(n+1)$ -th network of the sequence would be generated by adding one extra agent to the n -th network. Notice that, every countably infinite network can be obtained as the limit of such a sequence. Henceforth, with a slight abuse of notation, we write $n \rightarrow \infty$.

Before we begin our analysis we need two definitions. First, we say that *an action is used by a non-negligible share of the population* if the ratio between the number of agents choosing this action and the size of the population does not vanish to zero, as n becomes arbitrarily large. Formally, if $I_k^t = \{i \in N : a_i^t = \alpha_k\}$ is the set of agents who choose α_k at period t as $n \rightarrow \infty$, then:²⁶

²⁶One can think of I_k^t as the limit as $n \rightarrow \infty$ of a sequence of sets $\{I_{k,n}^t\}_{n=1}^\infty = \{\{i \in \{1, \dots, n\} : a_i^t = \alpha_k\}\}_{n=1}^\infty$, which is omitted for notational simplification.

Definition 1. α_k is chosen by a non-negligible share of the population at period t , if $\lim_{n \rightarrow \infty} \frac{|I_k^t|}{n} > 0$.

Second, we say that *an agent has unbounded neighborhood*, or equivalently that *an agent is connected to a non-negligible share of the population*, if there exists at least one agent that the ratio between the size of her neighborhood and the size of the population does not vanish to zero as n becomes arbitrarily large. Formally, if $|N_i|$ is called the *degree* of agent i and denotes the number of agents that i is connected with, then:

Definition 2. Agent i has unbounded neighborhood (is connected to a non-negligible share of the population), if $\lim_{n \rightarrow \infty} \frac{|N_i|}{n} > 0$.

Notice that, this is equivalent saying that for all $K \in \mathbb{N}$ it holds that $|N_i| > K$.

At first glance, one could doubt whether there is a difference between the cases of finite and arbitrarily large networks. Throughout this section we show, why the two cases are indeed different. Intuitively, we expect different behavior, since for arbitrarily large networks there must exist actions chosen by a non-negligible share of the population and for each of these actions, the probability of disappearance in finite time vanishes to zero.

Moreover, the possibility that some agents may be connected with a non-negligible share of the population makes them really crucial for the long-term behavior of the society. This is a property that changes the results of our analysis, even between different networks with arbitrarily large population. For this reason, we introduce the following assumption. (Keep in mind that the following assumption is used only when it is clearly stated.)

Assumption 1 (Bounded Neighborhood). *Exists $K > 0 \in \mathbb{N} : |N_i| \leq K$, for all $i \in N$*

Notice that, Assumption 1 does not hold if there exists an agent who affects a non-negligible share of the population. In the rest of the section, we compare the cases where Assumption 1 holds or does not hold, while stressing the conditions that make the results of Section 3 fail when the population becomes arbitrarily large.

1.4.1 Bounded Neighborhoods

In this part we assume that Assumption 1 holds. The main importance of this assumption is that it removes the cases where an agent can affect the behavior of a non-negligible share of the population.

For an arbitrarily large network, an obvious, nevertheless crucial, remark is that there must be at least one action chosen by a non-negligible share of the population. Our analysis will be different when there is exactly one such action and when there are more.

Proposition 2. *Consider any connected network with arbitrarily large population ($n \rightarrow \infty$), where there is only one action, α_k , initially chosen by a non-negligible share of the population. If the agents behave under imitate-the-best rule and Assumption 1 holds, then the network will converge with probability one to the monomorphic steady state, where every agent chooses α_k .*

The result of the above proposition is quite intuitive. If an action is chosen by a non-negligible share of the population and also each agent can affect finitely many others, then the probability that this action will disappear in a finite number of periods vanishes to zero. However, this is not the case for the rest of the actions, which face a positive probability of disappearance. Hence, no other action can survive in the long run. Nevertheless, Proposition 2 covers only the cases where there exists only one such action.

The question that follows naturally is whether the network has the same properties when more than one actions are diffused to a non-negligible share of the population. For a general network and payoff structure, the answer is negative. The negative result is supported by a counter example. More formally, consider a network with arbitrarily large population behaving under "imitate-the-best" rule and having bounded neighborhoods. Then, if there are more than one actions chosen by a non-negligible share of the population, we cannot ensure convergence to a monomorphic steady state, without imposing further restrictions to the network or/and payoff structure.

Example 2. Consider the following arbitrarily large network (Figure 1.2- Linear Network).

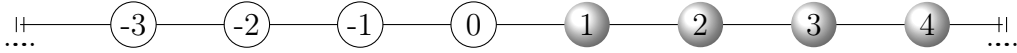


Figure 1.2: Linear Network

For simplicity, assume that the aggregate shock for each action is always equal to zero, i.e. $U_{AC}(a, z) = 0$, for all $a \in A$ and $z \in Z$. Therefore, the payoff is determined completely by the function U_{ID} . Notice that all the agents have bounded neighborhoods, since they are connected with exactly two agents and the diameter of the network, $d_N \rightarrow \infty$ as $n \rightarrow \infty$. In period t , there are two actions still present, α_1 and α_2 . Half of the population chooses each action; all the agents located from zero and to the left choose α_1 and the rest choose α_2 . Both actions have the same support of utilities, $U_{ID}(\alpha_1, \Omega) = U_{ID}(\alpha_2, \Omega) \in [0, 1]$. The probability density functions of the payoffs are as shown in the following graph.

For these distributions, an agent choosing action α_1 is, ex-ante, equally likely to receive higher or lower payoff compared to an agent choosing action α_2 , and vice-versa.²⁷ Moreover, the only agents who may change their choice are those located in the boundary between the groups using each action, i.e. agents 0 and 1 in period $t = 1$. This boundary will be moving randomly in the form of a random walk without drift. With probability $\frac{1}{4}(\frac{1}{4})$ it will be moving one step to the left (right) and with probability $\frac{1}{2}$ it will be staying at the same position. By standard properties of random walks, this boundary will never diverge to infinity, therefore the network will never converge to a monomorphic steady state. In fact it will keep fluctuating around zero, without ever reaching a steady state.

²⁷ $P[X_1 > X_2] = \int_0^1 P[X_1 > x_2] f_2(x_2) dx_2 = \int_0^{1/2} (1 - x_2)(\frac{1}{2} + 2x_2) dx_2 + \int_{1/2}^1 (1 - x_2)(\frac{5}{2} - 2x_2) dx_2 = \frac{1}{2}$

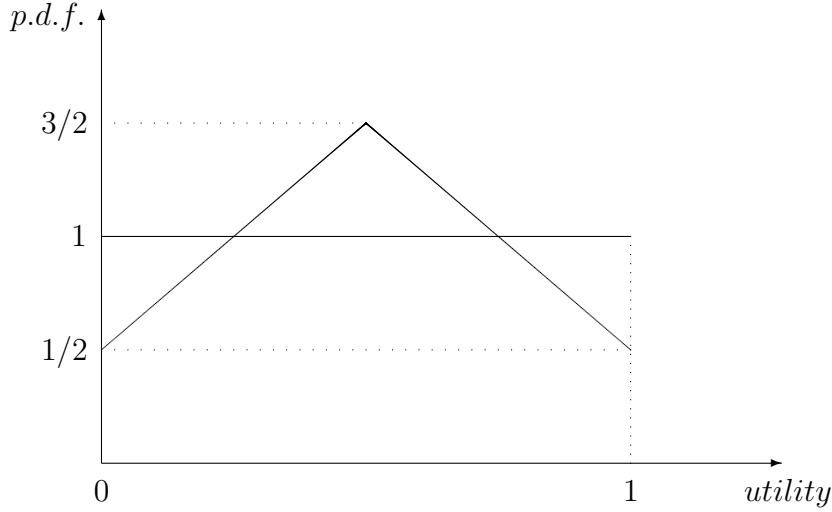


Figure 1.3: Probability density functions of utilities for α_1 and α_2 .

The negative result of the previous example does not allow us to ensure convergence to a steady state under every network structure, when at least two actions are chosen by a non-negligible share of the population. However, there exist sufficient conditions, related to the payoff and network structure, which can ensure it.

In the following proposition, we consider the case where all the agents have bounded neighborhoods and that all the remaining actions are chosen by non-negligible shares of the population. Under these conditions, we show that an action can be diffused to the whole population as long as it is sufficiently more likely to be successful than any other action. We focus on the case where there is no aggregate shock for each action ($U_{AC}(a, z) = 0$ for all $a \in A$ and for all $z \in Z$), in order to facilitate the exposition of the results. Nevertheless, the result for the general case would be totally analogous. Namely,

Proposition 3. *Consider a network with arbitrarily large population, with agents behaving under imitate-the-best rule and having bounded neighborhoods, with upper bound D . Moreover, each of the remaining actions, $\{\alpha_1, \dots, \alpha_m\}$ is chosen by a non-negligible share of the population, $\{s_1, \dots, s_m\}$ and $U_{AC}(a, z) = 0$ for all $a \in A$ and for all $z \in Z$. If there exists action α_k such that:*

- (i) $F_{ID}^{\alpha_k}(u) \leq [F_{ID}^{\alpha_{k'}}(u)]^D$, for all $k' \neq k$, and
- (ii) s_k is sufficiently large,

then the network will converge with probability one to a monomorphic steady state where every agent chooses action α_k .

Notice that D is an exponent of F .

Our sufficient condition is stronger than first order stochastic dominance; nevertheless we have the advantage of providing a result adequate for every network structure. The important fact in our proof, is that the agents changing to α_k are, in expectation, more than those changing from α_k to some other action. In general, this condition may depend not only on the payoff structure and the

initial share, but also on the network structure, which is something we completely disregarded in the present analysis. Nevertheless, this result can become the benchmark for future research on stronger conditions for specific structures.

It is somehow surprising that a condition as strong as first order stochastic dominance may not be sufficient. This happens either because of the complexity of the possible network structures, or because of insufficient initial share of the action of interest. To clarify this argument, we construct the following example.

Example 3. Consider the linear network described in Example 2. Initially, there are two actions present in the network, α_1 and α_2 . Similarly to the previous example, in the first period half of the population chooses action α_1 (all the agent located at zero and to the left) and, analogously, the rest choose α_2 . Notice that each agent has a neighborhood consisting of two agents apart from herself. Moreover, every period, there are only two agents, one choosing each action, who face positive probability of changing their chosen action.

Once again we assume that the aggregate shocks for each action are equal to zero. Therefore the payoffs depend only on the individual shocks. The cumulative distribution functions of $U_{ID}(\alpha_1, \cdot)$ and $U_{ID}(\alpha_2, \cdot)$ are such that $F_{ID}^{\alpha_1} < F_{ID}^{\alpha_2}$ (i.e. action α_1 First Order Stochastically Dominates action α_2), but $[F_{ID}^{\alpha_2}]^2 \leq F_{ID}^{\alpha_1}$.²⁸ Calling r_t the number of agents who change from action α_2 to action α_1 we get the following:

$$r_t = \begin{cases} 1 & \text{with probability } p_1 \\ 0 & \text{with probability } p_2 \\ -1 & \text{with probability } p_3 \end{cases}$$

FOSD ensures that $p_1 > p_3$. However, the second condition means that $p_1 \leq p_2 + p_3$, or equivalently $p_1 - p_3 < \frac{1}{2}$. The expected value of r_t will be $E[r_t] = (p_1 - p_3)$. So, $\frac{\lim_{\tau \rightarrow \infty} \sum_{t=1}^{\tau} r_t + s_1}{\lim_{n \rightarrow \infty} n} = p_1 - p_3 + s_1$. In order to get diffusion of α_1 we need $p_1 - p_3 + \frac{1}{2} \geq 1$, that cannot be the case here. Hence, although α_1 is FOSD compared to α_2 , α_2 will be chosen by non-negligible share of the population, even in the long run.

On the contrary, notice that if $[F_{ID}^{\alpha_2}]^2 \geq F_{ID}^{\alpha_1}$, then it can be the case (not necessarily) that $p_1 - p_3 \geq \frac{1}{2}$ and action α_1 is diffused to the whole population.

1.4.2 Unbounded Neighborhoods

In this section we drop the assumption of neighborhoods being bounded. This means that as n grows, there exists at least one agent whose neighborhood grows as well without bounds. If neighborhoods are unbounded, we cannot ensure convergence to a monomorphic steady state, even when the diameter of the network is finite. This happens because a single agent can affect a non-negligible share of the population, meaning that in one period an action can be spread from a finite number of agents

²⁸For example, let $F_{ID}^{\alpha_1}(u) = u$ and $F_{ID}^{\alpha_2}(u) = u^{\frac{2}{3}}$

to a non-negligible share of the population. This means that we will have more than one action chosen by a non-negligible share of the population, which may lead the network not to converge to a monomorphic state. We clarify the above statement with the following example (Example 4). Moreover, in case of unbounded neighborhoods, it does not hold anymore the result of Proposition 1, based on which the steady state has to be necessarily monomorphic. We provide another example where a network is in steady state and there are two actions present (Example 5).

Example 4. Consider the star network (Figure 1.4 - One Star), which satisfies finite diameter. As the population grows, the neighborhood of agent 1 also grows, without bounds.

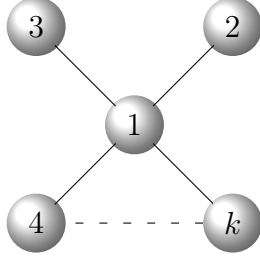


Figure 1.4: One Star

Suppose there are two actions present in the network, with same payoffs as in the Example 2 (no aggregate shocks for each action). Initially, all the peripheral agents choose action α_1 , while the central agent chooses α_2 . It is apparent that, as $n \rightarrow \infty$ the probability that the central agent will change her action in the second period will go to 1, because as n becomes arbitrarily large, then for any realized payoff of the central agent there will be at least one peripheral agent whose payoff will be higher than hers. However, reversing the argument, as n becomes arbitrarily large the probability that every peripheral agent will receive higher payoff than the central agent goes to 0. Then, for any realized payoff of the central agent, there will be a non-negligible share of peripheral agents who received lower payoff than hers. Hence, those agents will change to α_2 . This happens because $\Pr[U_{ID}(\alpha_1, \cdot) \leq \hat{u}] = F_{ID}^{\alpha_1}(\hat{u}) > 0$ for any $\hat{u} \in S_{ID}$.

A similar argument holds in every period, leading the choice of the central agent to change randomly from the second period on. This leads the network to a continuous fluctuation, where a non-negligible share of the population chooses each action in every period. Obviously, the system will never converge to a steady state.

We have shown that, without restricting the neighborhoods to be bounded we cannot ensure convergence to a steady state. In the following example, we show that when the neighborhoods are unbounded a network can be in steady state without this being monomorphic. This is in contrast to Proposition 1, where we have shown that for finite population, steady states are necessarily monomorphic. For arbitrarily large population, this need not be the case.

Example 5. Consider the following network (Figure 1.5 - Two Stars), that satisfies finite diameter and two of the agents have unbounded neighborhoods.



Figure 1.5: Two Stars

There are two actions present in the network, with same payoffs as in Example 2 (no aggregate shock for each action). All the agents on the left star of the figure (including the center) choose α_1 and similarly all the agents on the right star choose α_2 . As n grows, the neighborhoods of the central agents grow without bounds. They get connected with more and more agents choosing the same action as them and only one choosing differently. Hence, the probability that they will change action goes to 0, as n goes to infinity, hence they will continue acting the same, with probability one.²⁹ The peripheral agents have only one neighbor each, who always chooses the same action as them; so none of them will change her action either. Concluding, this network will be in a steady state where half of the agents choose each action.

The intuitive conclusion of this observation is that groups of agents who choose the same action and communicate almost exclusively between them can ensure the survival of this action in the long run, since no agent of this group ever changes her choice.

1.5 Conclusion

We study a model of observational learning in social networks, where agents imitate myopically the behavior of their most successful neighbor. We focus on identifying if an action can spread to the whole population, as well as the conditions under which this is possible. Our analysis reveals the different properties between finite and arbitrarily large populations.

For networks with finite population, we show that the network necessarily converges to a steady state, and this steady state has to be monomorphic. However, we cannot ensure which action will be the one to survive. This reveals the vulnerability of small populations to misguidance that can lead to the diffusion of sub-optimal actions. Furthermore, an action's performance in the initial periods, is crucial for its survival, since a sequence of negative shocks in early periods can lead to early disappearance of the action from the population. Moreover, despite the fact that we focus on connected networks, our analysis would be identical for connected components of disconnected networks. In that case, different actions would survive, one per each component. This raises the

²⁹As n grows, the extreme case where one of the two centers changes action will occur only if the other center receives the absolute maximum, but by continuity of the distributions of payoffs this will occur with probability 0.

issue that a potential explanation for the survival of different actions in different parts of societies may be due to lack of communication between them.

The results differ when the population is arbitrarily large. Under the assumption of bounded neighborhoods, an action is necessarily diffused, regardless of its efficiency, only when it is the only one to be chosen by a non-negligible share of the population. When more such actions exist, we provide a sufficient condition in the payoff structure, which can ensure convergence regardless of the network architecture. The general idea behind this result is that the diffusion of an action in a very large network is quite hard and requires either very special structure, very large proportion of initial adopters, or an action to yield higher payoffs much more often than the others.

Our sufficient condition is valuable mostly in networks where the largest neighborhood has quite small size. This increases the importance of studying the role of network architecture, rather than the payoff structure, when applying to networks where agents have large neighborhoods. The advanced complexity of this problem, makes it hard to deal with its general version. A very interesting and natural extension of the present paper would be to study different payoff structures in specific networks with importance in applied problems.

The role of central agents becomes apparent when we drop the assumption of bounded neighborhoods. Even an action initially chosen by a single agent can survive in the long run, if this agent is observed by a non-negligible share of the population. This stresses the influential power of massively observed agents in a society. Affecting the decision of a very well connected agent can become very beneficial for its diffusion.

In the present paper, we have studied some important properties of an "imitate-the-best" mechanism in a social network. However, crucial questions remain to be answered. Specifically, our analysis refers only to long-run behavior, without mentioning anything regarding neither the speed of convergence, nor the finite time behavior of the population. Different network structures are expected to have much different characteristics, leaving open space for future research.

1.6 Appendix

Proof of Proposition 1. Suppose that the statement does not hold. This means that at the steady state at least two different actions are chosen by some agent. For any structure, given that the network is connected, there will be at least a pair of neighbors, $i, j \in N$ choosing different actions. Moreover, given that the population is finite, each one of these agents must have a bounded neighborhood. Let $b_{i,1}$ and $b_{i,2}$ be the sets of neighbors of i choosing actions α_1 and α_2 respectively, with cardinalities $|b_{i,1}|$ and $|b_{i,2}|$. If we are at a steady state, then it must be that agent i will never change her choice. However, because of the common support and continuity of payoff functions, as well as boundedness

of neighborhoods, it holds always that:

$$\begin{aligned}
& Pr[U_j^t > U_l^t, \text{ for some } j \in b_{i,2} \text{ and for all } l \in b_{i,1}] \geq \\
& \geq Pr[U_{ID}(\alpha_2, \omega_j^t) > U_{ID}(\alpha_1, \omega_l^t) \text{ and } U_{AC}(\alpha_2, z_{\alpha_2}^t) > U_{AC}(\alpha_1, z_{\alpha_1}^t), \text{ for some } j \in b_{i,2} \text{ and for all } l \in b_{i,1}] \geq \\
& \geq Pr[U_{ID}(\alpha_2, \omega_j^t) > \hat{u}_{ID} \text{ and } U_{ID}(\alpha_1, \omega_l^t) < \hat{u}_{ID} \text{ and } U_{AC}(\alpha_2, z_{\alpha_2}^t) > \hat{u}_{AC} \text{ and } U_{AC}(\alpha_1, z_{\alpha_1}^t) < \hat{u}_{AC}, \\
& \text{for some } \hat{u}_{ID} \in S_{ID}, \hat{u}_{AC} \in S_{AC}, j \in b_{i,2} \text{ and for all } l \in b_{i,1}] = \\
& = \left\{ 1 - [F_{ID}^{\alpha_2}(\hat{u}_{ID})]^{b_{i,2}} \right\} [F_{ID}^{\alpha_1}(\hat{u}_{ID})]^{b_{i,1}} [1 - F_{AC}^{\alpha_2}(\hat{u}_{AC})] F_{AC}^{\alpha_1}(\hat{u}_{AC}) \geq p > 0
\end{aligned} \tag{1.2}$$

Where $F_{ID}^{\alpha_k}$ (respectively $F_{AC}^{\alpha_k}$) is the cumulative distribution function of U_{ID} (respectively U_{AC}) associated with action α_k , for $k = 1, 2$. Therefore, after each period, agent i has strictly positive probability p of changing her action. The probability that she will never change, given that no other agent in the network does so, is zero, because:

$$\lim_{t \rightarrow \infty} \prod_{t=1}^{\infty} (1-p) = \lim_{t \rightarrow \infty} (1-p)^t = 0$$

This leads to contradiction of our initial argument. The analysis is identical under the presence of more than two actions. Concluding, we have shown that it is impossible for a connected network with finite population to be in a steady state when more than one actions are still present. \square

Proof of Lemma 1. We begin the proof with some necessary additional notation. Suppose that, in period t there are at least two actions observed in the network. Consider the set which contains all players $i \in N$ who chose action α_k in period t , i.e. $a_i^t = \alpha_k$ and denote by $I_k^t = \{i \in N : a_i^t = \alpha_k\}$. Analogously, $\neg I_k^t = \{i \in N : a_i^t \neq \alpha_k\}$ is the set of agents who did not choose that action. Also, $F_k^t = \{i \in I_k^t : N_i \cap \neg I_k^t \neq \emptyset\}$ the set of agents who belong in I_k^t and who have, at least, one neighbor choosing an action different than α_k . Moreover, consider the set that contains those agents who have neighbors contained in F_k^t , without themselves being members of it; formally $NF_k^t = \{i \in I_k^t : N_i \cap F_k^t \neq \emptyset \text{ and } i \notin F_k^t\}$. Consider the following events:

$$\hat{b}^t = \{(\omega_i^t, z_{a_i^t}^t) \in \Omega \times Z, \forall i \in N_{i, \alpha_k}^t = \{F_k^t \cup NF_k^t\} : U_{ID}(a_i^t, \omega_i^t) \leq \hat{u}_{ID}, U_{AC}(a_i^t, z_{a_i^t}^t) \leq \hat{u}_{AC}\}$$

$$\hat{B}^t = \{(\omega_j^t, z_{a_j^t}^t) \in \Omega \times Z, \forall j \in N_i \setminus N_{i, \alpha_k}^t, \text{ where } i \in N_{i, \alpha_k}^t \text{ and } k' \neq k : U_{ID}(a_j^t, \omega_j^t) > \hat{u}_{ID}, U_{AC}(a_j^t, z_{a_j^t}^t) > \hat{u}_{AC}\}$$

for some $\hat{u}_{ID} \in \text{Int}(S_{ID})$ and $\hat{u}_{AC} \in \text{Int}(S_{AC})$. This means that in period t all agents who belong to N_{i, α_k}^t receive payoff lower than a certain threshold and all their neighbors who choose different actions receive payoff higher than this threshold.

We can define a lower bound of the probability that all agents who belong to F_k^t change their choice in the following period and none of their neighbors who was choosing differently changes to α_k . We denote this event as C^t . In fact, $C^t \supset \{\hat{b}^t \cap \hat{B}^t\}$, because the intersection of \hat{b}^t and \hat{B}^t is a specific case included in the set C^t . Hence, $Pr(C^t) \geq Pr(\hat{b}^t \cap \hat{B}^t) = Pr(\hat{b}^t)Pr(\hat{B}^t)$. Notice that we impose independence between \hat{b}^t and \hat{B}^t . This holds because the realized payoffs are independent across agents who choose different actions. Moreover, because of the full support of \mathbb{P} over $\Omega \times Z$ and the continuity of U_{ID} and U_{AC} in Ω and Z respectively we can ensure that both events occur with strictly positive probability, say $Pr(\hat{b}^t) = b_t > 0$ and $Pr(\hat{B}^t) = B_t > 0$. Hence,

$$Pr(C^t) \geq Pr(\hat{b}^t \cap \hat{B}^t) = Pr(\hat{b}^t)Pr(\hat{B}^t) = b_t B_t > 0$$

Notice that, C^t and C^{t+1} are conditionally independent, because the realization or not of C^t does not give any extra information about the realization of C^{t+1} . This is because all the shocks are independent across periods. Let $D_{\alpha_k}^t$ denote the event that, starting from period t , action α_k will disappear in the next d_N periods. One of the possible histories that leads to disappearance of action α_k is the consecutive realization of the events $C^{t+\tau}$ for $\tau \geq 0$, until all the agents who were using α_k at t have changed their choice, while no other agent has adopted it. The number of periods needed depends on the structure of the network. More specifically it is at most equal to the diameter of the network, which cannot be greater than $n - 1$.

The formal reasoning that proves the above argument is as follows. After a sequence of realizations $\{C^t, C^{t+1}, \dots, C^{t+\tau-1}\}$ for τ consecutive periods, all the agents who, in period t , were contained in I_k^t and whose distances from F_k^t were at most $\tau - 1$, will not be contained in $I_k^{t+\tau}$.³⁰ Given that the above argument is true for any $\tau \geq 0$, then after a sequence of such realizations for d_N consecutive periods all the agents in distance at most $d_N - 1$ from F_k^t will not be contained in $I_k^{t+d_N}$. However, if the state is not monomorphic then the maximum distance between any node $i \in I_k^t$ and the set F_k^t is $d_N - 1$.³¹ Therefore, after d_N consecutive realizations of $C^{t+\tau}$ action α_k will have disappeared from the network.

Hence, we can construct a strictly positive lower bound for the probability of occurrence of the event $D_{\alpha_k}^t$:

$$\begin{aligned} Pr(D_{\alpha_k}^t) &\geq Pr\left(\bigcap_{\tau=0}^{d_N-1} C^{t+\tau}\right) = \left[\prod_{\tau=1}^{d_N-1} Pr(C^{t+\tau} \mid C^{t+\tau-1})\right] Pr(C^t) \geq \\ &\geq \prod_{\tau=0}^{d_N-1} b_{t+\tau} B_{t+\tau} = \dot{B} > 0. \end{aligned}$$

□

Proof of Lemma 2. Denote as K_t the following set of histories:

$$K_t = \{\text{at time } t, \text{ exactly } K \text{ actions have disappeared from the network}\}$$

It is enough to show that $Pr(K_{t+\tau} \mid (K-1)_t) > 0$ where $\tau < \infty$ (in fact, at most $\tau = d_N$). Using the definition of $D_{\alpha_k}^t$ from Lemma 1 we get:

³⁰We define the distance between a node i and a set of nodes J as the minimum distance between i and some node $j \in J$.

³¹Suppose this does not hold. Then all agents in F_k^t should have distance d_N from at least one agent $i \in I_k^t$. Choose an $f \in F_k^t$, then, by definition of F_k^t , agent f should be connected with at least one agent $j \in \neg I_k^t$. However, agent j is connected with agents who have distance at least d_N from agent i , which means that the distance between agents i and j should be at least equal to $d_N + 1$, which is impossible by the definition of diameter d_N . Hence the initial argument holds.

$$Pr[K_{t+\tau} | (K-1)_t] = \sum_{\alpha_k \in A_t} \{Pr[D_{\alpha_k}^{t+d_N-1} | (K-1)_t]\} - Pr\left[\bigcup_{m=1}^{M-K-1} (K+m)_{t+\tau} | (K-1)_t\right] \quad (1.3)$$

where M is the total number of possible actions. Namely, the above expression tells that the probability of exactly one more action disappearing in the next τ periods equals the sum of probabilities of disappearance of each action, subtracting the probability that more than one them will disappear in the given time period. Analogously:

$$Pr[(K+1)_{t+\tau} | (K-1)_t] = \sum_{\alpha_k, \alpha_{k'} \in A_t}^{k \neq k'} \{P[D_{\alpha_k}^{t+d_N-1} \cap D_{\alpha_{k'}}^{t+d_N-1} | (K-1)_t]\} - Pr\left[\bigcup_{m=2}^{M-K-1} (K+m)_{t+\tau} | (K-1)_t\right]$$

$$Pr[(K+1)_{t+\tau} | (K-1)_t] + Pr\left[\bigcup_{m=2}^{M-K-1} (K+m)_{t+\tau} | (K-1)_t\right] = \sum_{\alpha_k, \alpha_{k'} \in A_t}^{k \neq k'} Pr[D_{\alpha_k}^{t+d_N-1} \cap D_{\alpha_{k'}}^{t+d_N-1} | (K-1)_t]$$

But:

$$Pr[(K+1)_{t+\tau} | (K-1)_t] + Pr\left[\bigcup_{m=2}^{M-K-1} (K+m)_{t+\tau} | (K-1)_t\right] = Pr\left[\bigcup_{m=2}^{M-K-1} (K+m)_{t+\tau} | (K-1)_t\right]$$

$$\Rightarrow Pr\left[\bigcup_{m=2}^{M-K-1} (K+m)_{t+\tau} | (K-1)_t\right] = \sum_{\alpha_k, \alpha_{k'} \in A_t}^{k \neq k'} Pr[D_{\alpha_k}^{t+d_N-1} \cap D_{\alpha_{k'}}^{t+d_N-1} | (K-1)_t]$$

Hence, by equation (2):

$$Pr[K_{t+\tau} | (K-1)_t] = \sum_{\alpha_k \in A_t} \{Pr[D_{\alpha_k}^{t+d_N-1} | (K-1)_t]\} - \sum_{\alpha_k, \alpha_{k'} \in A_t}^{k \neq k'} \{Pr[D_{\alpha_k}^{t+d_N-1} \cap D_{\alpha_{k'}}^{t+d_N-1} | (K-1)_t]\}$$

By Lemma 1, we have shown that the first summation is strictly larger than zero, and we just need to show that it is also strictly larger than the second one. Notice that the first summation is weakly larger trivially, because for any events A, B, C it holds that $Pr[A | C] \geq Pr[A \cap B | C]$. If the equality was possible, this would mean that the disappearance of any action would lead necessarily to the disappearance of another action. However, this cannot be the case because of the independence of payoff realizations across agents who choose different actions. Hence we have shown that: $Pr[K_{t+\tau} | (K-1)_t] \geq \ddot{B} > 0$, which completes our argument. \square

Proof of Lemma 3. The event of “convergence to a monomorphic state” is the same as telling that $M-1$ actions will have disappeared after T periods. One possible way of this to happen is if one action disappears every $\tau = d_N$ periods. Namely:

$$\{(M-1)_{t+T}\} \supset \{ \{(K)_{t+\tau} | (K-1)_t\} \cap \{(K+1)_{t+2\tau} | (K)_{t+\tau}\} \cap \dots \cap \{(M-1)_{t+T} | (M-2)_{t+T-\tau}\} \}$$

However, notice that the event on the right hand side of the expression consists of other independent events, because the states of nature are independent across periods. Hence, recalling the result of Lemma 2, we construct the following relation.

$$\begin{aligned} P_t[(M-1)_{t+T}] &\geq P_t[\{(K)_{t+\tau} \mid (K-1)_t\} \cap \{(K+1)_{t+2\tau} \mid (K)_{t+\tau}\} \cap \dots \cap \{(M-1)_{t+T} \mid (M-2)_{t+T-\tau}\}] \\ &= P_t[(K)_{t+\tau} \mid (K-1)_t] P_{t+\tau}[(K+1)_{t+2\tau} \mid (K)_{t+\tau}] \dots P_{t+T-\tau}[(M-1)_{t+T} \mid (M-2)_{t+T-\tau}] \\ &\geq \ddot{B}^{(M-K-1)} = \underline{C} > 0 \end{aligned}$$

This means that the probability of convergence in finite time is strictly positive. Moreover, notice that, for all $K < M-1 \Rightarrow \underline{C} < 1$. This remark is trivial because if $K = M-1$, the system has already converged, nevertheless we will use this to prove the following theorem. \square

Proof of Theorem 1. Lemma 3 shows us that the probability that the network will NOT converge to a monomorphic steady state in the next T periods, given that it has not converged until the current period, is bounded below 1. Formally:

$$P_t\{[(M-1)_{t+T}]^c \mid [(M-1)_t]^c\} \leq 1 - \underline{C} < 1$$

The event that the network will never converge is just the intersection of all the events where the network does not converge after $t+T$ periods, given that it has not converged until period t .

$$\{\text{The Network never Converges}\} = \{NC\} = \bigcap_{i=0}^{\infty} \{[(M-1)_{t+(i+1)T}]^c \mid [(M-1)_{t+iT}]^c\}$$

However, again these events are independent. Namely, the event of no convergence until $t+2T$ given no convergence until $t+T$ is independent of the event of no convergence at $t+T$ given no convergence until t : $\{[(M-1)_{t+2T}]^c \mid [(M-1)_{t+T}]^c\} \perp \{[(M-1)_{t+T}]^c \mid [(M-1)_t]^c\}$.

Hence, we can transform the above expression in terms of probabilities:

$$\begin{aligned} P[\{NC\}] &= \lim_{s \rightarrow \infty} \prod_{i=0}^s P\{[(M-1)_{t+(i+1)T}]^c \mid [(M-1)_{t+iT}]^c\} \\ &\leq \lim_{s \rightarrow \infty} (1 - \underline{C})^s = 0 \end{aligned}$$

So, the network will converge with probability one to a monomorphic steady state. \square

Proof of Proposition 2. Let $I_k^t = \{i \in N : a_i^t = \alpha_k\}$ be the set of agents who choose α_k at time t and analogously $\neg I_k^t = \{i \in N : a_i^t \neq \alpha_k\}$ the set of agents who do not. By construction of the problem, it holds that $\lim_{n \rightarrow \infty} \frac{|I_k^t|}{n} > 0$, hence $|I_k^t| \rightarrow \infty$ as $n \rightarrow \infty$. On the other hand, $\lim_{n \rightarrow \infty} \frac{|\neg I_k^t|}{n} = 0$, hence only finitely many agents do not choose α_k . For the rest of the notation recall Lemma 1.

Notice that, every action $k' \neq k$ is chosen by a finite number of agents, so the longest distance, $L_{k'}$, between an agent choosing k' and the closest agent choosing $k'' \neq k'$ must have finite length, $l \leq L_{k'}$. Hence, for all $k' \neq k$, the result of Lemma 1 still holds, if we substitute the diameter d_N by the maximum of all these distances, say L .

$$Pr(D_{\alpha_{k'}}^t) \geq Pr(\bigcap_{\tau=0}^L C^{t+\tau}) = \left[\prod_{\tau=1}^L Pr(C^{t+\tau} \mid C^{t+\tau-1}) \right] Pr(C^t) \geq \prod_{\tau=0}^L b_{t+\tau} B_{t+\tau} = \dot{B}' > 0$$

Notice, as well, the importance of the assumption for bounded neighborhoods. If this did not hold, then we could not ensure that the above product would be strictly positive. Moreover, the expression does not hold for the action α_k . The bounded neighborhoods' assumption can hold only if the diameter of the network grows without bounds as n grows. Given that α_k is the only action chosen by non-negligible share of the population, L_k must also grow without bounds. Hence, for any finite number of periods τ there exists n large enough, such that action α_k faces a zero probability of disappearance before period $t + \tau$. More intuitively, this happens because in each period, all the agents choosing a different action combined can affect the choice only of a finite number of other agents. Therefore, it is not possible that a non-negligible share of the population will stop choosing α_k in a finite time period.

Subsequently, the result of Lemma 2 still holds, with some appropriate modification. Namely $\tau = L$ and $Pr[K_{t+\tau} \mid (K-1)_t, I_k^{t+\tau} \neq \emptyset] \geq \ddot{B}' > 0$, meaning that action α_k cannot disappear from the network.

With similar reasoning, we get the modification of Lemma 3, which tells that there is positive probability of convergence to a monomorphic steady state in finite time, given that action α_k will not disappear. But, this is equivalent to the case where every agent chooses action α_k , denoted as $\{CA_k\}$. Formally, $P_t[\{CA_k\}] \equiv P_t[(M-1)_{t+T} \mid I_k^{t+T} \neq \emptyset] \geq \ddot{B}'^{(M-K-2)} = \underline{C}' > 0$.

Finally, as in Theorem 1, we get a similar expression, showing that the probability that agents' behavior will not converge to the same action, given also that action α_k is still present, it is bounded below 1. However, this is equivalent to the event of "Not converging to action α_k ". This is because action α_k cannot disappear, hence if the system converges to a single action, this has to be α_k . Namely:

$$\begin{aligned} P[\{NCA_k\}] &\equiv P_t\{[(M-1)_{t+T}]^c \mid [(M-1)_t]^c, I_k^{t+T} \neq \emptyset\} \leq 1 - \underline{C}' < 1 \\ P[\{NCA_k\}] &= \lim_{s \rightarrow \infty} \prod_{i=0}^s P\{[(M-1)_{t+(i+1)T}]^c \mid [(M-1)_{t+iT}]^c, I_k^{t+T} \neq \emptyset\} \\ &\leq \lim_{s \rightarrow \infty} (1 - \underline{C}')^s = 0 \end{aligned}$$

This means that the network will converge with probability one to a monomorphic steady state, where every agent will be choosing action α_k . \square

Proof of Proposition 3. Recall that we have restricted our attention to cases where there are only individual shocks, hence $U_{AC} = 0$ for all $a \in A$ and all $z \in Z$. Therefore the payoff is determined completely by U_{ID} . Notice that the following equivalence holds: $F_{ID}^{\alpha_k}(u) \leq [F_{ID}^{\alpha_{k'}}(u)]^D \Leftrightarrow Pr[U_{ID}(\alpha_k, \Omega) \geq \hat{u}] \geq 1 - Pr[U_{ID}(\alpha_{k'}, \Omega) \leq \hat{u}]^D$, for all $k' \neq k$ and $\hat{u} \in S_{ID}$. Before convergence occurs, there is at least one pair of agents such that $a_i = \alpha_k$ and $a_j = \alpha_{k'}$, with $k' \neq k$. Condition (i) ensures that for all the agents who observe at least one neighbor (including themselves) choosing α_k at period t (i.e. $\forall i \in N$ such that, $\exists j \in N_i : a_j^t = \alpha_k$), it holds that $Pr(a_i^{t+1} = \alpha_k) = p_1 > \frac{1}{2}$. Since

this holds for each agent we can construct the following variable, which we assume, in the first part of the proof, to be a finite number. Let:

$$r_{t+1} = \{\#i \in N : a_i^{t+1} = \alpha_k, a_i^t = \alpha_{k'}, k' \neq k\} - \{\#j \in N : a_j^{t+1} = \alpha_{k'}, a_j^t = \alpha_k, k' \neq k\}$$

In words, this represents the difference between the number of agents changing from any action $\alpha_{k'}$ to α_k , minus those that change from α_k to any $\alpha_{k'}$. The expected value of r_t in each period t will be:

$$E_n[r_t] = \sum_{i: \exists j \in N_i}^{a_j^t = \alpha_k} [p_1 - (1 - p_1)] = \sum_{i: \exists j \in N_i}^{a_j^t = \alpha_k} [2p_1 - 1] \geq r > 0, \text{ for all } t$$

The last inequality is direct implication of condition (i) and the assumption about finiteness of r_t . Since for every agent it is more probable to change to α_k , we expect that more agents will change from other actions towards α_k than the opposite.

In order to prove convergence, we need to show that in the long run and as the population grows it holds that:

$$\sum_t r_t + s_k n \rightarrow n \Leftrightarrow \lim_{n \rightarrow \infty} \frac{\lim_{t \rightarrow \infty} \sum_t r_t}{n} + s_k \geq 1 \quad (1.4)$$

By the weak law of large numbers, applied for the population, we get with probability one:

$$\lim_{t \rightarrow \infty} \sum_{\tau=1}^t r_\tau = \lim_{t \rightarrow \infty} \sum_{\tau=1}^t E_n[r_t] \geq \lim_{t \rightarrow \infty} r t$$

Which makes equation (6) equal to:

$$\lim_{n \rightarrow \infty} \frac{\lim_{t \rightarrow \infty} \sum_t r_t}{n} + s_k \geq \frac{\lim_{t \rightarrow \infty} r t}{\lim_{n \rightarrow \infty} n} + s_k = r + s_k \geq 1$$

The equality holds because $\lim_{t \rightarrow \infty} t \equiv \lim_{n \rightarrow \infty} n$. Finally, the last inequality comes from condition (ii) that requires s_k to be sufficiently large. Sufficiency depends on r , which itself depends on the distributions of the payoffs. Obviously, the more likely is action α_k to yield higher payoff than the other actions, the smaller initial population share is needed in order to ensure convergence.

Recall that we have assumed r to be a finite number. If r is a share of the population, it is apparent that the result still holds. In fact, it holds without any restriction on s_k , because the limit we calculated diverges to infinity. \square

Chapter 2 is dedicated to my brother Elias
and my parents Roula and Thomas

Chapter 2

Diffusion by Imitation: The Importance of Targeting Agents

2.1 Introduction

2.1.1 Motivation

The importance of social interactions for the diffusion of innovations, ideas and behavior is a topic that has attracted a lot of research interest over the years (see Jackson, 2008). Recent technological advances have made the collection and analysis of data related to the structure of relationships inside societies, as well as the rules guiding the behavior of their members possible. The appropriate use of this information can provide helpful tools for the effective diffusion of products, technologies and ideas in societies.

In this paper, we describe the optimal intervention of an interested party (from now on called *planner*) who seeks to maximize the diffusion of a given action in a society where agents imitate successful past behavior of their neighbors. Effective design of social influence campaigns appears to be a crucial problem in several real-life situations, usually when the planner has limited resources available. The direct example that comes to ones mind is the promotion of a new product by a firm. For instance, consider a firm that produces a new computer software and wants to establish it in a new market. It is rather common for people to seek advice from previous users before purchasing such. Moreover, firms tend invite advertise the software initially to a limited number of people that will then spread the word to the rest of the society. The correct selection of initial targets might be crucial for the success or not of the product. Furthermore, the effectiveness of this selection is affected by several factors, such as the horizon in which the firm expects to observe the outcome of the targeting strategies, or how the initial targets are going to be distributed around the society.

There are several other examples one could think that fit the general idea of our work. For instance, consider an NGO or a government that wishes to promote a new highly productive or environmentally friendly agricultural technology and for doing so it wants to select a few farmers in one or several villages to adopt the technology initially. Or else, a government that wishes to reduce

criminal activity and is willing to sponsor a number of ex-criminals to change their lifestyle. Or even, a political or religious organization that wishes to propagate its ideology and needs to locate a number of initial seeds in the society to spread the word to their neighborhood. As one can see, the problem of optimal social influence is directly applicable to a bunch of different environments and seemingly unrelated areas.

Of course, in order to obtain tractable and intuitive results we need to make a set of simplifying assumptions, which in some sense might reduce the applicability of our analysis to certain problems. Nevertheless, we provide a framework that can help us understand better some parameters that affect crucially problems of social influence and we illustrate how beneficial the knowledge about society's network structure may be for the efficient design of marketing and general social influence strategies.¹

There are several reasons why economic agents adopt simple behavioral rules, such as the imitation of successful past behavior. For example, they often need to make decisions without knowing the potential gains or losses of their possible choices. Furthermore, when these situations arise with high frequency and the agents' computational capabilities are limited, then they tend to rely on information received from past experience of others, rather than experimenting themselves.² These arguments are also supported by a recent, but growing, empirical and experimental literature which provides strong evidence in favor of the fact that in several decision problems the agents tend to imitate those who have been particularly successful (see Apesteguia et al., 2007; Conley and Udry, 2010; Bigoni and Fort, 2013).

Most of the existing literature on targeting in networks has focused on the importance of central agents (see Ballester et al., 2006). Having a high or a low number of connections (see Galeotti and Goyal, 2009; Chatterjee and Dutta, 2011), or diffusing information to many others who are poorly connected (see Galeotti et al., 2011) are some usual characteristics of influential agents. The importance of these characteristics is obvious and beyond doubt. Nevertheless, we show that there is another important factor with significant effect on diffusion. This is whether the targeted agents are concentrated all together, in the sense that they are connected between them, or they are spread around the network. Notice that, in order to use this additional tool, information about the exact structure of the network is required. Each of these strategies may be optimal depending on certain parameters, with one of the most important factors being planner's patience.

Throughout our analysis we highlight the differences between the optimal targeting strategies of an *infinitely patient planner*, i.e. one who cares about the diffusion of her preferred action in the long run, and an *impatient planner*, i.e. one who cares about the diffusion of her preferred action in the short run. To our knowledge, this is the first paper that determines both short and long run optimal targeting strategies and provides a clear distinction between them. Interestingly enough, we find the optimal targeting strategies to differ sharply in the two cases and this difference to persist

¹In broader terms, the importance of social networks in efficient marketing design has been studied extensively and in several different disciplines. For an extensive list see Galeotti and Goyal (2009) and references therein.

²These are some of the reasons why imitation has been subject to extensive theoretical study in different environments (see Ellison and Fudenberg, 1993, 1995; Vega-Redondo, 1997; Eshel et al., 1998; Schlag, 1998; Alós-Ferrer and Weidenholzer, 2008; Duersch et al., 2012).

for all values of the other parameters. This comparison is important in several realistic scenarios, since different targeting strategies may be appropriate depending on the time horizon.

2.1.2 Setting and Results

Formally, we consider a finite population of behaviorally homogeneous agents located around a circle. In each period, all agents choose simultaneously between two alternative actions. The stage payoff each action yields is uncertain and depends on a random shock, which is common for all the agents who have chosen the same action in that period.³ Shocks are independent across actions and across periods. There are no strategic interactions between agents. After making their decisions, all agents observe the chosen actions and the realized payoffs of their two immediate neighbors. Subsequently, they update their choice myopically, imitating the action that yielded the highest payoff within their neighborhood in the preceding period. Notice that, an agent who does not observe any of her neighbors choosing the alternative action never changes her choice.

Another problem that fits our model well is that of the diffusion of agricultural technologies.⁴ Farmers' harvests depend mostly on common factors such as the weather and the fertility of the land. Moreover, it is normal to assume that farmers are aware of the technologies and crops used by their neighbors, as well as the payoffs they receive. In particular, Conley and Udry (2010) show that farmers tend to imitate those who have been very successful in the past, whereas as it is pointed out by Ellison and Fudenberg (1993) the farmers' technology decisions are guided mostly by short-term considerations, especially when capital markets are poorly developed or malfunctioning.

The planner is interested in maximizing the diffusion of her preferred action in the population. She can be either infinitely patient, therefore interested in the diffusion of the action in the long run; or impatient, therefore interested in the diffusion of the action after just one period. She is assumed to know the structure of the network, as well as how agents behave and she can intervene in society by enforcing a change at the initial choice of a subset of the population. Ideally, she would like to target the whole population, but doing so in reality would be extremely costly. Hence, our goal is to identify the planner's optimal targeting strategy given the number of agents she is able to target.

Observe that all the agents are identical with respect to any measure of centrality. In fact, none of them has any positional advantage or disadvantage compared to the rest of the population. Despite this fact, we find that expected diffusion changes substantially depending on the subset of the population that has initially been targeted by the planner. This highlights how important it is for the planner to know the exact structure of the network.

We show that the optimal targeting strategy depends on two parameters: (i) the likelihood of the planner's preferred action being more successful than its alternative and (ii) the planner's patience. In fact, we observe a sharp contrast between the optimal strategies of an infinitely patient planner and that of a very impatient one. More specifically, when the planner's preferred action has

³This assumption is made only in order to facilitate the tractability of the results.

⁴See also Ellison and Fudenberg (1993).

higher probability of being more successful than its alternative, then the optimal targeting strategy for an infinitely patient planner is to concentrate all the targeted agents in one connected group;⁵ whereas when this probability is lower it is optimal to spread them uniformly around the network. Interestingly, for a very impatient planner, the optimal targeting strategy is exactly the opposite.

The intuition is relatively simple and depends on the fact that in the long run only one of the two actions survives. Therefore, when the action is likely to be successful, then an infinitely patient planner wants to prevent its disappearance due to a few consecutive negative shocks in the first periods. For this reason she prefers to concentrate them all together. To the contrary, if the action is unlikely to be successful, then the optimal strategy for the planner is to try and take advantage of a possible sequence of successful shocks during the first periods. By concentrating all targeted agents together, she would only manage to make her preferred action disappear more slowly, since for its diffusion a large number of consecutive successful shocks would be needed, which is rather unlikely to happen.

For an impatient planner the arguments are reversed. When the preferred action is likely to be successful, then the planner wants to make it visible to as many agents as possible, therefore she should spread the initial adopters⁶ around the society. On the other hand, if the action is more likely to be unsuccessful, then the planner wants to prevent as many of the initial adopters as possible from observing the alternative action, therefore she should concentrate them all together.

At this stage, one could question how important is the role of the particular behavioral rule for obtaining these results and how robust they would be under either Bayesian (see Gale and Kariv, 2003) or De Groot (see DeGroot, 1974; Golub and Jackson, 2010) learning rules. A crucial aspect for the current results is that the agents do not accumulate information through time, which has two negative and one partially positive effect. On the one hand, the society is vulnerable to misguidance by certain unexpected events even at later stages, which for example should not be the case if the agents perform Bayesian learning. Moreover, under any initial conditions there is no guarantee that the society would converge to the planner’s desired action (even if this is the socially optimal). On the other hand, the process is less path dependent than De Groot learning, where initial opinions may drive a society towards an inefficient state even with certainty. Once again, this would not be a problem under Bayesian updating, which however has been acknowledged by the literature of learning in networks to require extremely complex calculations.

We extend our analysis in many different directions. We discuss the optimal strategies of planners with intermediate levels of patience, thus intending to identify how the transition between the two extreme cases occurs. Moreover, we quantify the practical meaning of infinite patience by characterizing the expected waiting time before convergence occurs. We observe that, for those cases in which the planner’s optimal strategy is to concentrate all the initial adopters in one group the process is slowed down substantially. In addition to this, we discuss what happens if we allow for inertia and

⁵In the circle network, a connected group is a segment of the circle.

⁶Throughout the paper we interchange between the terms targeted agents and initial adopters, since the targeted agents are those who adopt the action in the first period.

show that the results remain unchanged. This extension captures many realistic features, such as the existence of switching costs and some forms of conformity. Finally, we repeat our analysis for the linear and the star network and provide numerical simulations for other network structures, as an attempt to identify the effect of centrality on our results.

2.1.3 Related Literature

The role of influential agents in social networks has been studied in different disciplines, such as computer science (see Kempe et al., 2003; Richardson and Domingos, 2002) and marketing (see Kirby and Marsden, 2006), as well as in economics. Intuitively, a crucial feature is the centrality of an agent, which depicts either the number of immediate neighbors an agent has or how important she is for the connectivity of the network (see Ballester et al., 2006).

Other environments similar to ours have been studied in physics, mathematics and computer science, especially in the areas of cellular automata and voter models. The most similar paper to ours is Bagnoli et al. (2001), who study the long-run behavior and the phase transition of a system with characteristics similar to ours. A crucial difference with our paper is that they assume initial conditions to be random, given that they refer to initial positions of particles. Hence, they do not focus on identifying optimal initial conditions, which is the main focus of the present paper.

Similar models have also been used in the Economics literature. In particular, Ortuño (1993) considers a standard voter model setting, where agents are located in a two dimensional infinite lattice and probabilistically adopt the choice made by one of their neighbors. This is the only article where centrality does not play a role. The agents are homogeneous in preferences and location and there exists a planner who seeks the diffusion of a technology in the society and can choose between targeting a single connected segment of agents, or spread them uniformly around the population. The author concludes that these two choices lead to the same probability of diffusion; result which is guided mainly by the infinite size of the lattice. In our paper, the population is finite and we do not restrict the potential choices of the planner. This features allow us to obtain more general results. More recently, Yildiz et al. (2011) generalize the standard voter model by introducing “stubborn” agents, i.e. who never change their choice. Similarly to us, the authors also discuss the problem of optimal placement of stubborn agents, when trying to maximize their impact on the long run expected choices of agents.

Our paper is closely related to Galeotti and Goyal (2009) and Goyal and Kearns (2012). The research questions are similar, however both papers study different types of processes and focus on different aspects. Galeotti and Goyal (2009) focus on “word-of-mouth” communication and on social conformism, which are mechanisms that disregard the performance of a product. They look for the optimal influence strategy of a firm who seeks to maximize the diffusion of its product in a society show that, depending on the learning mechanism, agents with low or high number of connections should be targeted, underlining again the important role of centrality. Apart from analyzing different learning mechanisms, there are two more main differences with our work. First,

the focus is completely on the short run optimal targeting strategy, using a two-period model. Second, in their case the network is modeled as a degree distribution, where each agent meets a fixed number of agents every period, who are randomly drawn from a population mass. Therefore, the probability that two agents have a common friend is zero. This approach disregards potential information about the exact structure of the network, which we show to be important for the planner. Moreover, another question that remains unanswered is what would happen if the degree distribution tends to a uniform and furthermore whether we could do better by knowing the exact formation of the network. Both of these questions can be tackled in our environment with promising results.

Along similar lines is the work of Goyal and Kearns (2012), where the authors study competition between two firms who distribute their resources trying to maximize the adoption of a product by consumers located in a social network. The modeled process is significantly different from ours and the paper focuses more on how efficiently the resources are allocated, as well as on the effect of budget asymmetries on the equilibrium allocations.

In a recent paper, Galeotti et al. (2011) show that in a setting where information transmission is strategic (in contrast to what happens here) influential agents are those who diffuse information to many others, who themselves are poorly connected. Another closely related paper is by Chatterjee and Dutta (2011). The authors study the optimal behavior of a firm that seeks to diffuse a technology in a society considering different network structures, focusing mostly on the linear and the circular network. The fact that part of the population consists of perfectly rational agents and that the structure of the network is common knowledge changes significantly the dynamics of the system. Also, the firm is allowed to target a single agent. They find that a firm that produces a good quality product should target the agent with maximum decay centrality,⁷ whereas a firm that produces a bad quality product should target the agent with the highest number of connections. There is an apparent analogy between these results and ours, in the sense that in both cases the optimal choice depends on the quality of the technology, which in our case could be loosely approximated by the probability of the action of interest being more successful.

Regarding agents' behavior, we focus on imitation of successful past behavior. There is a recent but growing empirical literature (see Apesteguia et al., 2007; Conley and Udry, 2010; Bigoni and Fort, 2013) that provides empirical evidence on the adoption of this behavior in real environments. In theoretical framework, Vega-Redondo (1997) has shown that a Cournot economy, where the agents follow this rule, converges to the Walrasian equilibrium. Moreover, Alós-Ferrer and Weidenholzer (2008), Eshel et al. (1998) and Fosco and Mengel (2011) study coordination games and public good games respectively, played between neighbors, where the agents imitate their most successful neighbor. Their focus is mostly on the characterization of stochastically stable configurations.

In a general framework, the current analysis builds upon the work on learning from neighbors (see Banerjee, 1992; Banerjee and Fudenberg, 2004; Bala and Goyal, 1998; Chatterjee and Xu, 2004;

⁷Consider a decay parameter δ , where $0 < \delta < 1$, then decay centrality measures the proximity between a given node and every other node weighted by the decay. Formally, this is equal to $\sum_{j \neq i} \delta^{n(i,j)}$, where $n(i,j)$ is the geodesic distance between nodes i and j .

Ellison and Fudenberg, 1993, 1995; Gale and Kariv, 2003). These articles study different learning mechanisms in environments where the agents face common individual problems and there are no strategic interactions. They mainly discuss conditions under which efficient actions spread to the whole population. In particular, Ellison and Fudenberg (1993) use an environment very similar to ours, in the sense that the agents repeatedly choose between two alternative technologies whose payoff depends on a random shock, which is common for all the agents who use the same technology. Apart from these similarities in the setting, there are several differences regarding the role of the network and other details of the model, but the main difference is that their focus is not on the characterization of the optimal intervention in favor of one of the two technologies.

The rest of the paper is organized as follows. In section 2.2 we formally define the model. Section 2.3 contains the characterization of the optimal targeting strategy for an impatient planner, whereas in Section 2.4 we analyze the optimal targeting strategy for an infinitely patient planner. In Section 2.5 we briefly discuss some extensions and Section 2.6 concludes. A thorough study of the extensions can be found in Appendix A. All proofs can be found in Appendix B.

2.2 The Model

2.2.1 The Agents

There is a finite set of agents $N = \{1, \dots, n\}$, referred to as *population* of the network. At time $\tau = 1, 2, \dots$, each agent $i \in N$ chooses between two alternative actions, $a_i^\tau \in \{A, B\}$. Each action yields random payoff. The payoff of agent i is independent of the other agents' choices. Therefore, interactions among agents are not strategic and their connections represent only an exchange of information. Moreover, agents who choose the same action at a given period receive equal payoffs.⁸ The payoffs of both actions change in each period, with the realizations being independent across periods. Action B yields strictly higher payoff than action A with probability $p \in (0, 1)$, while action A yields strictly higher payoff than B with probability q . For the derivation of the main results, we focus on the case where $q = 1 - p$. Notice that, for $q = 1 - p$, the probability of both actions yielding exactly equal payoffs is zero, which is assumed only in order to avoid unnecessary tie-breaks and does not affect the results. In the extensions section we relax this assumption and allow for different values of q .⁹ The ratio between p and q turns out to be crucial for our analysis, therefore, we define $r = \frac{p}{q}$. From now on, we will say that there is a *success* (*failure*) in period τ if action B (A) yielded higher payoff in this period.¹⁰

⁸This assumption does not affect the main intuitions and is imposed mainly in order to facilitate the tractability of the results.

⁹Different values of q allow us to introduce the possibility of inertia. Later on, we use this feature to capture scenarios where either the two actions can have the same realized payoffs, or there are externalities between the actions of one's neighbors, or switching costs.

¹⁰Later on, with some abuse of terminology, we will also define the success and failure in a random walk, in a similar way.

The *planner* is an agent, outside of the population, who seeks to maximize the diffusion of action B in the population. She can do so by changing to her favor the choice of a subset of the population before the beginning of the first period.¹¹ Throughout the paper, these are mentioned as *targeted agents*. Optimally, she would like to affect the whole population, but in reality this would be extremely costly. This cost enters implicitly if we assume that the cardinality of the subset she can affect is fixed exogenously.¹² More specifically, given that at period $\tau = 0$ all the agents are choosing action A , the planner can target $t \leq n$ agents from the population and make them choose B in period $\tau = 1$. After that, the planner cannot affect the society anymore. The goal of this paper is to characterize the planner's optimal targeting strategy.

The planner can be either impatient or infinitely patient. A planner is called *impatient* if she cares about the diffusion of her preferred action after only one period. Similarly, a planner is called *infinitely patient* if she cares only about the diffusion of her preferred action in the long-run.¹³ We find that the optimal behavior of an impatient planner is exactly the opposite to that of an infinitely patient planner. In the Extensions section, we also discuss some intermediate levels of planner's patience.

2.2.2 The Network

A social network is represented by a family of sets $\mathcal{N} := \{N_i \subseteq N \mid i = 1, \dots, n\}$, when N_i is called i 's neighborhood and denotes the set of agents observed by agent i . We assume N_i to contain i as well. In the main part of our analysis, we examine the *circular* network, where each agent interacts with her two immediate neighbors, i.e. $N_i = \{i-1, i, i+1\}$ for $i = 2, \dots, n-1$, whereas $N_1 = \{n, 1, 2\}$ and $N_n = \{n-1, n, 1\}$. The current structure imposes an undirected network, because for all $i, j \in N$, $j \in N_i$ if and only if $i \in N_j$. In this setting, the network structure describes the flow of information in the network, in the sense that each agent $i \in N$ can observe the action and the realized payoff of her neighbors, $j \in N_i$.

2.2.3 The Behavior

At the end of each period, the agents observe the actions and realized payoffs of their neighbors. Subsequently, they have the opportunity to revise their choices. Revisions happen simultaneously

¹¹Assuming that the planner intervenes only once is a simplifying assumption. Nevertheless, if the planner could target agents in multiple periods, then all one would have to do is to repeat the same analysis multiple times. Intuitively, multiple periods of targeting could allow the planner to be more risky in the beginning, in the sense of targeting larger number of groups, thus intending to attract the whole society quickly and then condition her future targeting strategy on the realized history. In a slightly different environment, where the planner does not know perfectly the likelihood of success of each action, multiple periods of targeting could help the planner improve the accuracy of her belief over this. Such an explanation would be more plausible for environments such as the one presented in Tsakas (2014).

¹²In the section of extensions in Appendix A we endogenize the number of affected agents and we discuss the returns to investment for different values of it.

¹³Throughout the paper, by diffusion of an action we refer to the expected number of adopters over a certain number of periods, either one or infinite. Implicitly, this means that the planner is assumed to be risk neutral.

for all agents.¹⁴ We assume that each agent $i \in N$ can observe the choices and the realized payoffs of her neighbors in the previous period. According to these observations, she revises her choice by imitating the most successful action within her neighborhood in the preceding period. Notice that, an agent never switches to an action that she did not observe, i.e. that neither her nor any of her neighbors chose in the previous period. Moreover, if an action disappears from the population, then it never reappears.

The important aspect of this myopic behavior is that the agents discard most of the information available. They ignore whatever has happened before the previous period, hence they are unable to form beliefs about the underlying payoff distributions of their alternative choices.

2.2.4 The Problem

After the planner has chosen the initial adopters, the population consists of s agents choosing action A (from now on called *non-adopters*) and t agents choosing action B (from now on called *adopters*); obviously $s + t = n$. We call a *group* a sequence of neighboring agents all of whom choose the same action and are surrounded by agents choosing the opposite action. The population is formed of m groups of neighboring agents who choose action A , with sizes $\{s_1, s_2, \dots, s_m\}$, where $\sum_{k=1}^m s_k = s$ and analogously m groups of neighboring agents who all choose action B , with sizes $\{t_1, t_2, \dots, t_m\}$, where $\sum_{k=1}^m t_k = t$.¹⁵ We refer to these groups as groups of *type A* and of *type B* respectively. The numbering of the groups is based on their size in increasing order, $s_1 \leq s_2 \leq \dots \leq s_m$ and $t_1 \leq t_2 \leq \dots \leq t_m$. With some abuse of notation we also use s_1, s_2, \dots, s_m and t_1, t_2, \dots, t_m to name the groups.

Our goal is to find the optimal size of all s_k and t_k for $k = 1, \dots, m-1$,¹⁶ their optimal position (if it matters), as well as the optimal number of groups, m .

In order to avoid unnecessary complications in the calculations (which arise without the gain of any additional intuition) we assume that every group must have an *even* number of agents. The qualitative intuition of the results would be the same even without imposing this assumption, and in most of the cases it would not affect them at all. It is only imposed for better exposition of the results. We refer the reader to the analysis of the linear network (Appendix A) for a more detailed motivation regarding this assumption. Formally:

Assumption 2 (A1). s_i and t_i are even numbers for all $i \in \{1, \dots, m\}$.

¹⁴Simultaneous updating is assumed to keep the problem tractable. In the section of extensions, we provide a simple intuition of how this, as well as not perfectly correlated payoffs, would affect them quantitatively, but not qualitatively.

¹⁵Notice that, the fact that the network is a circle and there exist exactly two actions ensures that the number of groups is the same for both actions.

¹⁶ s_m and t_m must be equal to $s_m = s - \sum_{k=1}^{m-1} s_k$ and $t_m = t - \sum_{k=1}^{m-1} t_k$ respectively.

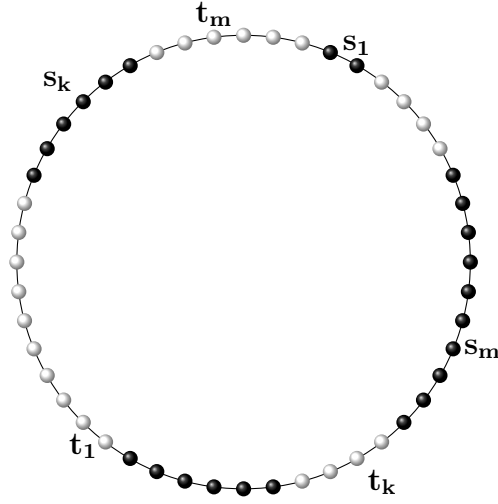


Figure 2.1: Example of an initial configuration: White nodes represent agents choosing action B and black nodes agents choosing action A .

2.3 Results for an Impatient Planner

In this section, we study the optimal targeting strategy of a planner who cares about maximizing the expected number of agents choosing action B after exactly one period. Figure 2.2 shows the two possible configurations after one period. White dots represent the agents who choose initially action B and black dots those agents who choose initially action A . Observe that only those agents who are on the boundary of a group can change their choice. In fact, for m denoting the total number of groups, in case action B is more successful in the first period, then there will be $2m$ additional adopters in the next period; whereas, in case action A is more successful, the number of adopters will decrease by $2m$. The probabilities of ending in each of the two possible states is p and $q = 1 - p$ respectively.



Figure 2.2: Initial configuration and the two possible configurations after one period.

Hence we can construct the objective function of the impatient planner:

$$EN_B(1) = t + 2mp - 2m(1 - p) = t + 2m(2p - 1)$$

It is easy to see that the optimal targeting strategy depends on the probability p of action B being more successful. Namely, for $p > 1/2$ the objective function is strictly increasing in m .¹⁷ Therefore, it is optimal to have as many groups as possible. On the other hand, if $p < 1/2$ the objective function is strictly decreasing in m and therefore it is optimal to locate all initial adopters in one group.¹⁸ This result is formally stated in the following proposition:

Proposition 4. *Under (A1), then for an impatient planner and for a given number of initial adopters, t*

- *If $p > 1/2$, the optimal targeting strategy is to spread the initial adopters to as many groups as possible, i.e.*
 - *If $t < s$, then $m = \frac{t}{2}$, with $t_1 = \dots = t_m = 2$.*
 - *If $t > s$, then $m = \frac{s}{2}$, with $s_1 = \dots = s_m = 2$.*
- *If $p < 1/2$, the optimal targeting strategy is to concentrate all the initial adopters in one group, i.e. $m = 1$, $t_1 = t$ and $s_1 = s$.*

Observe that, as long as the planner creates the maximum number of groups, the allocation of the agents inside these groups is not important. This feature will change slightly in the case of infinite patience.

Intuitively, this result suggests the following. If an action is likely to be successful, then the planner should try to make it directly visible to as many non-adopters as possible. In doing so, she will manage to attract the maximum number of additional adopters. On the contrary, if an action is unlikely to be successful, then the planner prefers to prevent most of the initial adopters from observing the opposite action. As a result, even upon an unsuccessful realization, most of the initial adopters will not observe the alternative action, therefore they will not revise their choice in the second period. As we will see, this optimal strategy changes sharply if the planner is infinitely patient.

2.4 Results for an Infinitely Patient Planner

In this section, we study the optimal behavior of an infinitely patient planner, i.e. one who cares only about the diffusion of her preferred action in the long run. A crucial feature of this setting is that such a planner disregards completely the speed of the procedure. Before beginning our analysis, it is useful to state two prior results.

¹⁷If $q \neq 1 - p$, we would simply have to replace $2p - 1$ with $p - q$.

¹⁸Throughout the paper, we disregard the case of $p = \frac{1}{2}$. This is because for $p = \frac{1}{2}$ and given Assumption 1 every targeting strategy of the planner yields exactly the same result.

2.4.1 Preliminaries

Diffusion when Agents Imitate-the-Best Neighbor

The present behavioral rule constitutes a special case of “Imitate-the-Best Neighbor” rule, applied in a setting of individual decision-making under uncertainty without strategic interactions between agents. Agents observe the choices of their neighbors and the payoff those choices yield. Subsequently, they revise their choices repeatedly according to these observations. In particular, they do so by imitating the action that yielded the highest payoff within their neighborhood in the preceding period.

In such a setting, it turns out that every connected network converges with probability one to a steady state where all the agents choose the same action (see Tsakas, 2013). Moreover, if $p \in (0, 1)$ then any of the actions can be the one to survive in the long run. This is based on the fact that all actions are vulnerable to a sequence of negative shocks, which can lead to their disappearance. Given that an action which disappears from the network never reappears, it turns out that only one of them survives at the end.

In our case, this result ensures that only one of the two actions will survive in the long run and that both actions have a positive probability to be the ones succeeding. Hence, the optimal strategy for an infinitely patient planner is the one that maximizes the probability that action B gets diffused to the whole population in the long run. We define this probability as follows:

Definition 3. $P_B(\cdot)$ is the probability that action B will be diffused to the whole population in the long run.

This probability will depend not only on the size of the population n , the number of targeted agents t and the probability of success p , but also on the choice of the planner about which agents to target. Notice as well that maximizing this probability is equivalent to maximizing the expected number of agents choosing action B in the long run; a remark that will clarify the analogy between our short run and long run analysis.

Results on Random Walks with Absorbing Barriers

A technical result which turns out to be particularly useful comes from Kemeny and Snell (1960). It refers to the properties of a finite one-dimensional random walk with absorbing barriers, which in the current context is defined as a Markov chain whose state space is given by the integers $j \in \{0, 1, \dots, n\}$ and its initial state is i . For some numbers p and q satisfying $0 < p, q < 1$, the transition probabilities are given by $P_{j,j+1} = p$, $P_{j,j-1} = q$ and $P_{j,j} = 1 - p - q$ for all $j \neq 0, n$ and $P_{0,0} = P_{n,n} = 1$. The endpoints of a random walk are called absorbing barriers because upon reaching one of them random walk eventually stays there forever. Those two states are the only absorbing ones. We denote the probability of absorption at state n (respectively at state 0) when the process initiates at state i , by $P_n(i)$ (respectively $P_0(i)$). Specifically, Kemeny and Snell (1960) compute the random walk’s probabilities of absorption at each one of the two absorbing states as follows:

Lemma 4 (Kemeny and Snell (1960)). *Consider a random walk with state space $\{0, 1, \dots, n\}$, where both barriers 0 and n are absorbing. If the probability of moving to the right (from j to $j+1$) is p , the probability of moving to the left (from j to $j-1$) is q , and $r = \frac{p}{q}$, then the probability of absorption at state n , when starting from state i is:*

$$P_n(i) = \begin{cases} \frac{r^n - r^{n-i}}{r^n - 1} & \text{if } p \neq q \text{ (or equivalently } r \neq 1) \\ \frac{i}{n} & \text{if } p = q \text{ (or equivalently } r = 1) \end{cases} \quad (2.1)$$

Analogously, the probability of absorption at the state 0 is $P_0(i) = 1 - P_n(i)$.

For the moment, $p \neq q$ is equivalent to $p \neq \frac{1}{2}$. Later on we extend our analysis to show that the results are completely analogous in the more general case.

To help us understand how this result can be used to express the current diffusion process we consider a linear network (see Figure 2.3), with agents named $\{1, 2, \dots, n\}$, where each agent has two neighbors, except of agents 1 and n who have one neighbor each. At period $\tau = 1$, agents 1 to i choose action B and the rest choose action A . Hence, every period only two agents may revise their choice (for example in the first period those are the agents i and $i+1$). The border fluctuates until either agent n chooses B , or agent 1 chooses A . The position of the right border of adopters follows a random walk with absorbing barriers 0 and n . Notice that, in order for agent 1 to choose A the left barrier must be located at the artificial node 0, which is going to be omitted in most of the graphs for expositional simplicity. Hence, we can use the result stated above to describe the probability of diffusion for each of the two actions. We call a random walk *successful* (*unsuccessful*) if it ends up in the absorbing state where all the nodes included in the walk choose action B (A).

This result is particularly helpful for our analysis, because any initial targeting choice induces a stochastic process that can be expressed as a sequence of conditionally independent random walks with absorbing barriers, similar to the one described above. Despite having multiple borders between groups, all of them fluctuate synchronously, because the payoffs for each action are perfectly correlated and therefore all agents on the boundaries make the same choice in each period. For example, considering the beginning of the process, each border fluctuates until either the smallest group of adopters, with size t_1 or the smallest group of non-adopters, with size s_1 disappears. This process can be represented by the random walk that is shown in Figure 2.6. Upon success or failure of the first walk, the process starts fluctuating according to a new random walk that depends on the smallest still existing groups of each type.

2.4.2 Main Results

Not surprisingly, the ratio $r = \frac{p}{q} = \frac{p}{1-p}$, which describes the likelihood of action B being more successful than action A , is a crucial parameter. However, surprisingly enough, this is the only parameter that affects the optimal targeting strategy and more specifically whether r is higher or lower than 1.¹⁹ The most interesting result, though, is that the optimal strategy of the infinitely

¹⁹Even though, at the moment $r > 1$ is identical to $p > 1/2$, we keep this notation because it facilitates the extension to cases where $q \neq 1 - p$.

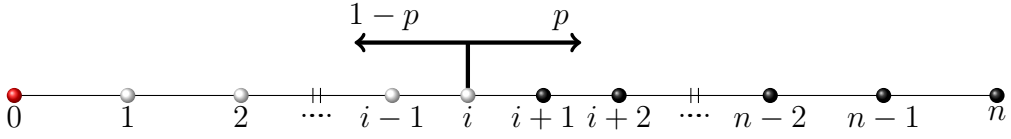


Figure 2.3: Initial configuration of a random walk with absorbing barriers 0 and n .

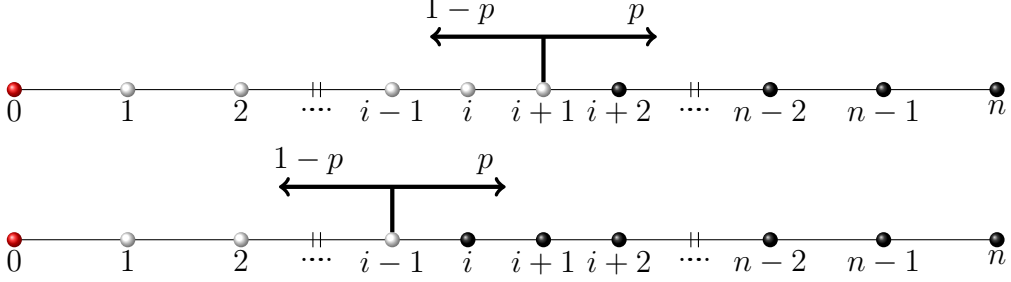


Figure 2.4: The two possible configurations after one period.

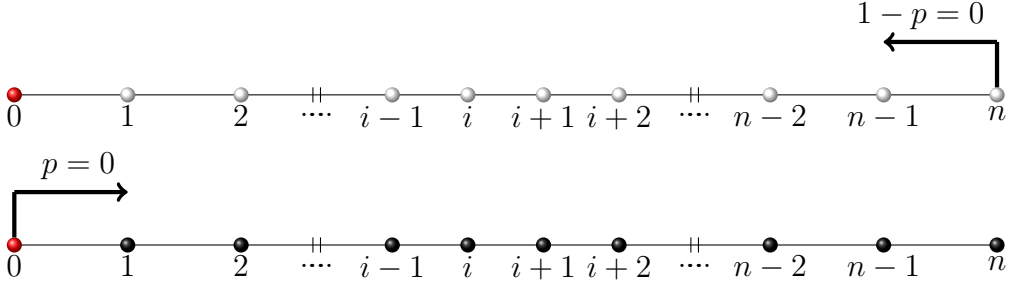


Figure 2.5: The two possible configurations after absorption.

patient planner is in complete contrast to that of the impatient planner. In particular, we observe that, for $r > 1$ the optimal targeting strategy of the planner is to concentrate all the targeted agents in a single group, whereas for $r < 1$, the optimal strategy is to spread them as much as possible across the population, splitting them into as many groups as possible and as symmetrically as possible. Observe that, the results are not just different, but are exactly opposite to those found for the impatient planner. In fact, the optimal strategy for an infinitely patient planner is the worst possible strategy for an impatient planner and vice versa.

For a better exposition of the general results, we split the problem into three sub-problems. First, we consider the case where the groups are restricted to be symmetric. Then, we consider the asymmetric case where the planner can target up to two groups and finally we consider the general asymmetric case.

Symmetric cases

First, we consider the symmetric case, where the groups of agents choosing the same action are restricted to have equal sizes, namely $s_1 = \dots = s_m = \frac{s}{m}$ and $t_1 = \dots = t_m = \frac{t}{m}$. Assuming no problems of divisibility we find the optimal number of groups, m . As we have mentioned already,

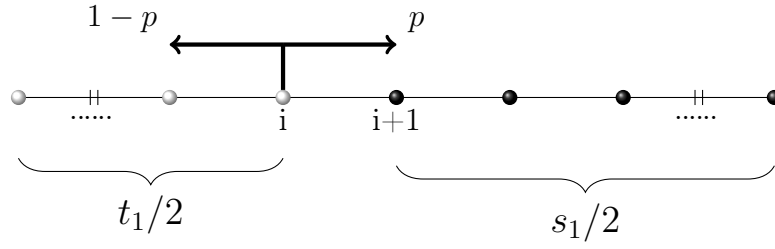


Figure 2.6: The first random walk

the optimal targeting strategy depends only on the ratio r . In particular, when $r > 1$ it is optimal to concentrate all initial adopters in one group, whereas when $r < 1$, it is optimal to split them in as many groups as possible, i.e. $m = \min\{s/2, t/2\}$. Formally:

Proposition 5. *Under (A1) and given $s_1 = \dots = s_m = \frac{s}{m}$ and $t_1 = \dots = t_m = \frac{t}{m}$, then for an infinitely patient planner:*

- If $r > 1$, then $\arg \max_m P_B(m|s, t, n, r) = 1$
- If $r < 1$, then $\arg \max_m P_B(m|s, t, n, r) = \min\{s/2, t/2\}$

All proofs can be found in Appendix B.

Intuitively, this proposition suggests that when the probability of success is high, it is beneficial to concentrate all initial adopters together. This prevents the disappearance of the preferred action upon the realization of a sequence of negative shocks during the first periods. The opposite strategy is optimal when the probability of success is low. Then the planner wants to take advantage of some potential good shocks during the first periods, which will spread the action to as many agents as possible.

Asymmetric Cases

We now turn attention towards the more general asymmetric cases. At first, consider the case where the planner is restricted to target at most two groups of each type, with sizes s_1, s_2 and t_1, t_2 respectively. Then the process can be described as shown in the Figure 2.7. Recall that $s_1 \leq s_2$ and $t_1 \leq t_2$.

It turns out that the optimal decision depends completely on the value of r . More specifically, if $r > 1$ it is optimal to concentrate all the agents in one group, while if $r < 1$ the optimal choice is to have two groups of equal size for each action.

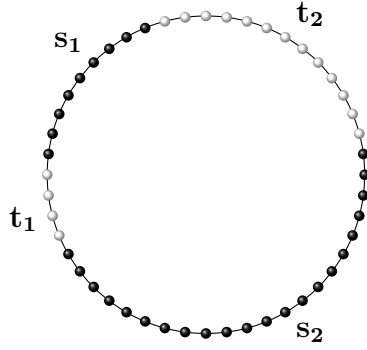


Figure 2.7: Example of an initial configuration with two groups of each type

Proposition 6. *Under (A1) and given $m \leq 2$, then for an infinitely patient planner:*

- *If $r > 1$, the optimal targeting strategy is to concentrate all initial adopters in one group.*
- *If $r < 1$, then the optimal targeting strategy is to split the initial adopters into two as equal as possible groups, and locate them in the population as symmetrically as possible, i.e. $s_2 - s_1 \leq 2$ and $t_2 - t_1 \leq 2$*

The intuition is similar to that of the symmetric case. However, an interesting finding is that this result does not hold for all restrictions on m . For example, if we restrict the number of groups to be not greater than three, then it is not optimal to split the agents into three equal groups of each type. Hence, it is not the case that we always prefer symmetric configurations compared to asymmetric ones. Notice that, this would be a sufficient condition for the proof of our main result, but it does not always hold. Nevertheless, this does not affect our general result which is stated below.

The two propositions help us construct the main theorem of the paper which describes the optimal targeting strategy in the general case of m initial groups of each type, allowing them to be of different sizes. The result is in line with the previous findings and suggests that when $r > 1$, then the optimal choice is to concentrate all the initial adopters in one group; whereas when $r < 1$, then the optimal choice is to spread them uniformly across the population in as many and as equal groups as possible. Namely:

Theorem 2. *Under (A1), for an infinitely patient planner*

- *If $r > 1$, the optimal targeting strategy is to concentrate all the initial adopters in one group, i.e. $t_m = t$ and $t_1 = \dots = t_{m-1} = 0$ for any m .*
- *If $r < 1$, the optimal targeting strategy is to spread the initial adopters in as many groups as possible and locate these groups as symmetrically as possible, i.e.*
 - *If $t < s$, then $m = \frac{t}{2}$, with $t_1 = \dots = t_m = 2$ and $s_m - s_1 \leq 2$,*
 - *If $t > s$, then $m = \frac{s}{2}$, with $s_1 = \dots = s_m = 2$ and $t_m - t_1 \leq 2$*

As it has been mentioned already, the importance of this result lies in the complete contrast between the optimal strategy of an infinitely patient planner in comparison to an impatient one. An infinitely patient planner prefers to protect from some initial negative shocks an action which is more likely to be successful from some initial negative shocks, whereas she prefers to spread as much as possible an action which is more likely to be unsuccessful, trying to take advantage of a few positive shocks in the first periods. When the probability of success is low, she knows that by concentrating all the targeted agents together, a lot of positive shocks will be needed in order to capture the whole population, which is rather improbable for an action that is expected to be often unsuccessful.

2.5 Extensions

In this section, we briefly present the results of several extensions that address specific questions related to the problem of interest. One can find an thorough discussion of these extensions in Appendix A.

A natural question that arises from the contrast between the optimal behavior of an impatient and an infinitely patient planner is what happens for intermediate levels of patience. First of all, we observe that the expected diffusion is very sensitive to small changes in the initial configuration and therefore it becomes particularly hard to construct a general strategy for all intermediate levels of patience. Nevertheless, we discuss the optimal targeting strategy of a planner who cares about the diffusion of action B after 3 periods and we obtain an enlightening result. Namely, if $r > 1$, then in some cases the planner prefers to spread the initial adopters in groups consisting of four, instead of two, agents. Whereas, if $r < 1$, she always needs to compare between the two extreme cases, i.e. concentrating all of them in one group or spread them to as many groups as possible. This result provides a useful starting point to understand how the optimal targeting strategy changes as the planner becomes more patient. We refer the reader as well to the extension that refers to the expected time before total diffusion occurs, we study an additional simple exercise that can be seen as an intermediate case of patience.

Furthermore, we discuss what happens if the number of targeted adopters, t , is endogenous. To do this it is necessary to define explicitly the profit function of the planner. We focus on linear cost of targeting and we find that for sufficiently low (high) unit cost the planner prefers to target all the (none of the) agents. Interestingly enough, for intermediate unit costs (with the bounds depending on the different parameters, but mostly on p) the planner prefers to target an intermediate number of agents, $0 < t^* < n$. This result is quite intuitive if one thinks that targeting additional agents increases the probability of total diffusion of action B , but never ensures it. Therefore, if a planner has targeted sufficiently many agents, then the additional expected benefit from targeting one more might not compensate the cost of targeting this agent.

Another crucial aspect we would like to shed more light on is the practical meaning of infinite patience. In reality, no planner can wait literally infinitely many periods. We try to identify the expected time before total diffusion of one action occurs and how the planner's strategies that maxi-

mize the expected diffusion affect this expected time. Not surprisingly, we find that a larger number of groups leads to faster diffusion and therefore the optimal strategy of the planner for $r > 1$ has the drawback of maximizing also the expected time before total diffusion occurs. On the other hand, for $r < 1$ the optimal strategy is also the one that leads to the fastest expected time of total diffusion. An interesting feature, which is in line with standard results in the analysis of random walks, is that the expected time of total diffusion explodes as r gets very close to 1. In general we find that for different configurations the expected time until diffusion occurs may vary substantially and therefore one must be very careful when acting as an infinitely patient planner.

In addition to this, one might argue that fast total diffusion might not be always optimal for the planner, because this might lead to a fast disappearance of her desired action from the population. For this reason, we perform a simple exercise where the planner obtains a positive gain from total diffusion of her preferred action and a negative one from total diffusion of the alternative action. We find that for extreme values (either high or low) of p the optimal strategies of the planner might be determined by the speed at which diffusion occurs, rather than by the probability of successful diffusion. In fact, irrespectively of the level of patience, if p is very low the planner prefers to locate all the agents together, whereas if it is very high prefers to spread them as much as possible. The intuition is similar to that described for the impatient planner. This can be also seen as an additional attempt to characterize optimal strategies for intermediate levels of patience, but should be approached with caution since it is only a partial result.

Moreover, we discuss cases where inertia is possible. This generalization allows us to capture some realistic scenarios, which include the possibility of both actions having equally good realizations, the existence of switching costs and some forms of conformity. Such settings can be captured by allowing $q \neq 1 - p$, where q is the probability of action A being more successful than action B , and we explain why our results are not affected by this feature.

A question similar to the possibility of inertia would be how the results would be affected either if the payoffs were not perfectly correlated in each period, or if updating was not occurring simultaneously. Intuitively, either of these two scenarios would induce more realizations of the random shocks, which in turn should favor quantitatively the action that is more likely to be successful. However, given that all the targeting strategies would be affected on the same way, the results should not be affected qualitatively. Nevertheless, one should be cautious when making such claims and a concrete answer would be possible only after a systematic analysis of these scenarios.

Additionally, we discuss the optimal targeting strategy of a planner for some slightly modified network structures, so as to get an idea of how the results would be affected by the presence of central agents. First, we study the linear network assuming that the planner can target a single group of agents. Once again, we find a sharp contrast between the optimal targeting strategies of an impatient and an infinitely patient planner. In particular, for an infinitely patient planner, if $r > 1$ it is optimal to target one of the two corners, whereas if $r < 1$ it is optimal to target the agents who are located around the center. To the contrary, for an impatient planner, if $r > 1$ it is optimal to target any segment of the line that does not include any of the corner agents, whereas if $r < 1$ it is

optimal to target one of the two corners. In this part of our analysis, we also drop the assumption of groups having an even number of agents and we show why dropping this assumption complicates our analysis without providing additional insights. We also discuss briefly the star network, in which the vast importance of very central agents becomes apparent.

Finally, we run a set of simulations to test the robustness of our results to the addition of a few links in the circle network and we find that our conclusions remain valid. In particular, if $r > 1$ it is always optimal to target only one group of connected agents, with the optimal location of the group depending on the position of the additional links. Conversely, if $r < 1$ it is almost always optimal to target the subset of agents that minimizes the number of successful draws needed to capture the whole population. This is in line with our previous findings and provides intuitions which can be useful for the study of more general network structures.

2.6 Conclusion

We have analyzed the optimal intervention of a planner who seeks to maximize the diffusion of an action in a circular network where the agents imitate successful past behavior of their neighbors. It turns out that there is room for strategic targeting of initial adopters even in environments where all agents are completely identical. Knowledge of the exact structure of the network, and not only its degree distribution, can be beneficial for a planner. We find that the optimal decision depends almost completely on two parameters. On the likelihood r of the preferred action being more successful and on how patient the planner is. Changes in these two parameters lead to completely opposite optimal behavior.

Assuming that the planner knows the exact structure of the network, as well as the location of each agent, might seem quite strong and tough to be satisfied in large networks. Nevertheless, it is exactly this assumption that allows us to focus on the importance of the targeted agents' relative positions, which in turn brings into consideration several new insights on the planner's problem. One can see this paper as a first step on understanding the effect of agents' exact positions in a network for certain diffusion processes and use it as a benchmark for problems where this assumption can be partially relaxed. A similar argument applies to the assumption regarding the particular behavioral rule, which turns attention from Bayesian (see Gale and Kariv, 2003) and DeGroot (see DeGroot, 1974; Golub and Jackson, 2010) learning rules towards more naive ones. A natural step one would consider is to construct targeting mechanisms that are robust to uncertainty over the exact network structure and behavioral rule. In particular, an interesting avenue for future research would be to identify targeting strategies that are able to perform well (are close to optimal) for a broad family of difference networks and under uncertainty over the rules that govern the behavior of the agents.

Moreover, throughout the paper we have disregarded completely in our analysis is the risk aversion of the planner. We have assumed the planner to be risk neutral, caring only about the expected number of adopters. For a risk averse planner, we would expect the optimal behavior to contain more dispersed targets than for the risk neutral one, but this remains an open question for future

research.

The current paper constitutes a first attempt to explore targeting possibilities in networks where agents imitate successful behavior. Given that our network structure is relatively simple, a natural extension would be to explore which of the current features are still present and which of these are changing when passing to more general network structures. It is apparent that centrality features arising in more complex networks will play an important role. However the exact characteristics remain to be studied.

2.7 Appendix A - Extensions

2.7.1 Intermediate patience - the τ -patient planner

The sharp contrast between optimal targeting strategies of impatient versus infinitely patient planner makes plausible the question of how this transition occurs as the planner becomes more patient. Intermediate levels of patience may be defined in several alternative ways, of which we choose the following:

Definition 4. *The planner is τ -patient if she cares about maximizing the expected number of agents choosing action B at period τ , denoted by $EN_B(\tau)$.*

Although, at first sight, this definition of impatience might seem unusual, it covers all important intuitions and at the same time it can be easily extended to more complicated cases. For example, the qualitative results would be very similar if the planner cared about the sum of discounted expected number of agents choosing her preferred action from period one up to period τ .

First of all, we analyze the case where the planner is 2-patient, i.e. that she cares only about the expected number of adopters after two periods. Notice that, after two periods, there are four possible states the society can be in.²⁰ Notice, also, that the number of groups may have decreased after the first period. This is because, after the first period, the groups consisting of only two agents who chose the action that was unsuccessful during that period will disappear. For a visual representation, look at the configuration occurring after failure in Figure 2.2, where the bottom left group of two agents choosing action B disappears after a failure for action B in the first period. Hence, we can construct easily the objective function of a 2-patient planner as:²¹

$$EN_B(2) = t + 2(2p - 1)\{2m - [p\alpha_2 + (1 - p)\beta_2]\}$$

²⁰In general after τ -periods there are 2^τ possible histories, so there are 2^τ possible configurations.

²¹To get this expression one should consider the four histories that are possible after two periods. If there is a success in the first period, there will be added $2m$ adopters and will disappear the α_2 groups of type A that consist of two agents. Hence, if there is a success (failure) in the second period, there will be added (subtracted) $2(m - \alpha_2)$ agents. We proceed analogously for the case of failure in the first period. Each of these histories has a probability of occurring. Therefore, the objective function will be equal to $EN_B(2) = t + p^2[2m + 2(m - \alpha_2)] + p(1 - p)[2m - 2(m - \alpha_2)] + p(1 - p)[-2m + 2(m - \beta_2)] + (1 - p)^2[-2m - 2(m - \beta_2)]$, which can be simplified to the following expression.

where α_2 is the number of groups consisting of two agents choosing A and β_2 is the number of groups consisting of two agents choosing action B . Notice that there are two opposing forces. We analyze them for $p > 1/2$, since for $p < 1/2$ the analysis is exactly the opposite. For $p > 1/2$, on the one hand, we would like to have as many groups as possible, since this increases the expected number of agents choosing B . On the other hand, we would like not to have groups of size two, since this enters in the objective function negatively. Hence, as long as we can create more groups, with size larger or equal than four agents, it is preferred to do so. Now, what would happen if in order to create one extra group, we would have to create one or more groups with size two, for each one of the actions. The most difficult condition to be satisfied arises if we are left with groups of size no larger than four agents for both actions. In this case, in order to create one additional group, we would need to substitute one group of four agents with two groups of two agents for each type. Therefore, we would need to increase the number of both α_2 and β_2 by two. The necessary condition such that we prefer to have one extra group would be:

$$2m - [p\alpha_2 + (1 - p)\beta_2] \leq 2(m + 1) - [p(\alpha_2 + 2) + (1 - p)(\beta_2 + 2)]$$

With some straightforward calculations, we see that this condition is satisfied always with equality, hence we are indifferent between having this one additional group or not. Given that this is the toughest condition to satisfy, this means that in all the other cases it is strictly preferred to have one more group. Concluding, for a 2-patient planner it is always preferred, at least weakly, to have one extra group. Therefore, her optimal behavior is identical to the one of the 1-patient planner.

The intuition is basically the same for the 3-patient planner. Notice that, during the first two periods it is possible that also groups of four agents disappear, which may affect the total number of groups in the third period. The objective function of the planner becomes:

$$EN_B(3) = t + 2(2p - 1)\{3m - 2[p\alpha_2 + (1 - p)\beta_2] - [p^2\alpha_4 + (1 - p)^2\beta_4]\}$$

where α_4 (β_4) is the number of groups consisting of four agents who choose action A (B) and α_2, β_2 are as defined above.

We observe that, in this case it would be preferable not to have groups consisting neither of two nor of four agents. Hence, there is an ambiguous trade-off between having more groups and those additional groups consisting of four or less agents. One can easily observe that the creation of groups consisting of four agents has very small negative effect, compared to the positive effect of the addition of a group. This means that the planner prefers to have two groups with four agents, rather than one with eight. However, there are still more cases to analyze. The case captured in Figure 2.8 is the one that imposes the toughest condition to satisfy. In particular, the question is whether the planner prefers to break two groups of four agents each into four groups of two agents each.

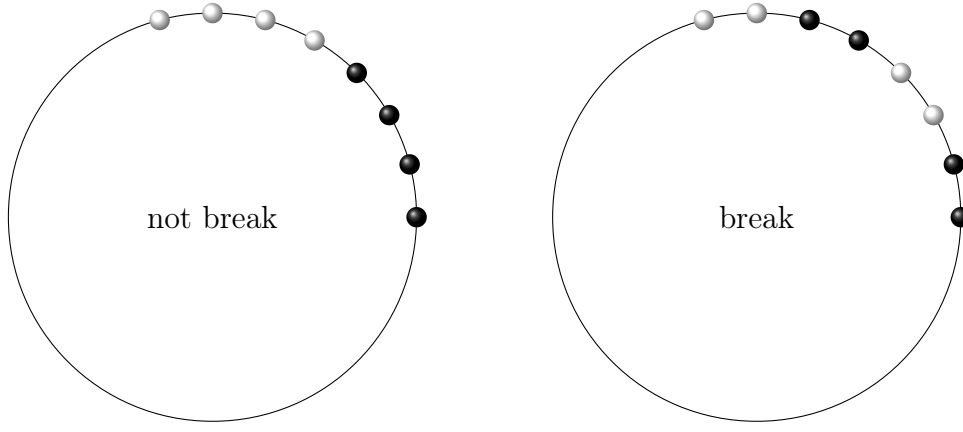


Figure 2.8: Configuration where a 3-patient planner prefers not to create an additional group if $r > 1$

The condition is the following and it is not satisfied for any p :

$$\begin{aligned}
& 3m - 2[p\alpha_2 + (1-p)\beta_2] - [p^2\alpha_4 + (1-p)^2\beta_4] \leq \\
& \leq 3(m+1) - 2[p(\alpha_2+2) + (1-p)(\beta_2+2)] - [p^2(\alpha_4-1) + (1-p)^2(\beta_4-1)] \Leftrightarrow \\
& \Leftrightarrow 3 - 4p - 4(1-p) + p^2 + (1-p)^2 \geq 0
\end{aligned}$$

In a similar manner, we need to check whether other possible ways of creating more groups are beneficial. This case turns up to be the only one where the 3-patient planner prefers not to have more groups. The second stricter condition arises by the configuration depicted in Figure 2.9. In this case, it turns out that the planner is indifferent between the two alternative configurations. All the other conditions are strictly satisfied.

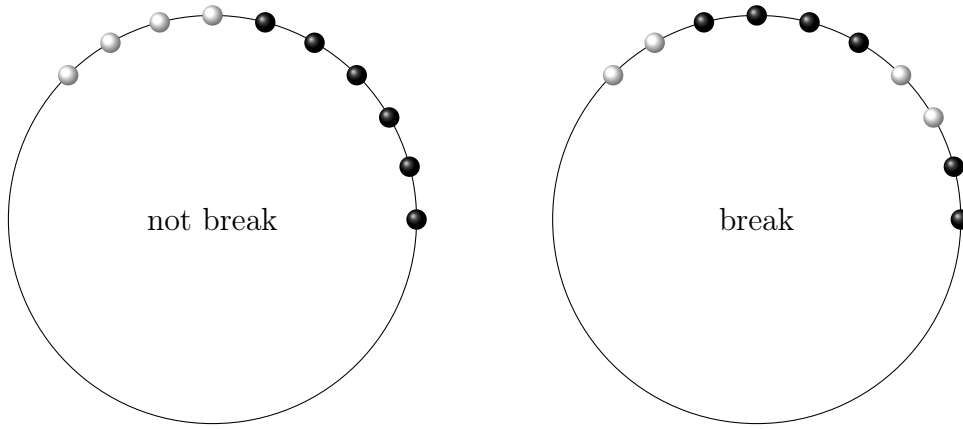


Figure 2.9: Configuration where a 3-patient planner is indifferent between creating or not an additional group, if $r > 1$.

We can describe the optimal behavior of a 3-patient planner as follows. For $p > 1/2$, if $t \lll s$ (if t is much smaller than s),²² then optimally $t_1 = t_2 = \dots = t_m = 2$ and $m = t/2$. For $s \lll t$,

²²We say that t is much smaller than s if $t < s/2$, because in this case even for the maximum number of $t/2$ groups,

optimally $s_1 = s_2 = \dots = s_m = 2$ and $m = s/2$. However, if $s \approx t$ (if s is approximately equal with t) then if t is a multiple of 4 the optimal targeting strategy is to choose $t_1 = t_2 = \dots = t_m = 4$ and $m = t/4$ and if t is not a multiple of 4, then $t_1 = 2$, $t_2 = \dots = t_m$ and $m = \frac{t+2}{4}$ and the same for s .

Analogously, for $p < 1/2$, the objective function of the planner decreases in the number of groups until $m = t/4$ and increases afterwards. Hence, we need to compare the cases where $m = 1$ and $m = \min\{s/2, t/2\}$. Comparing the two cases for $s \approx t$, we get the condition:

$$3 \geq 3\frac{t}{2} - 2p\frac{s}{2} - 2(1-p)\frac{t}{2}$$

which is satisfied whenever $t \geq 6$. Hence, it is not optimal to target one group only when $t \approx s$ and $t < 6$.

The case of the 3-patient planner provides very useful insights regarding the change in the planner's optimal behavior as she becomes more patient. On the one hand, for actions with high probability of success ($p > 1/2$) it seems that this transition happens smoothly, since it becomes optimal for the planner to target larger and larger groups. On the other hand, when the action has low probability of success ($p < 1/2$) this transition happens suddenly, meaning that for low values of t it is optimal to target as many groups as possible, whereas for high values of t it is optimal to target only one group.

Continuing the analysis for more patient planners would give more complicated results which would not be easily tractable. This is because the results would depend vastly on the exact initial position of each group and not only on their number and sizes. A complete characterization for any value of planner's patience would be an interesting and illuminating extension to our paper.

2.7.2 Endogenous number of targeted agents

In this section, we endogenize the number of targeted agents t and we discuss the returns to investment for both an impatient and an infinitely patient planner, focusing on the optimal targeting strategies we have already determined. An important first step is to observe the changes in the objective function of the planner, without taking into account the cost of targeting agents.

For an impatient planner the objective function is $EN_B = t + 2m(2p - 1)^{23}$, where for $p > 1/2$ it is optimal to target the maximum number of groups, which is $m^* = \min\{s/2, t/2\}$, recalling that $s = n - t$. Whereas, for $p < 1/2$ it is optimal to target one group. Therefore, it is easy to see that:

$$EN_B^*(t) = \begin{cases} 2pt & \text{if } p > 1/2 \text{ and } t \leq n/2 \\ 2(1-p)t + n(2p-1) & \text{if } p > 1/2 \text{ and } t > n/2 \\ t + 2(2p-1) & \text{if } p < 1/2 \end{cases}$$

which is obviously linearly increasing in t .

no groups of type A needs to consist of less than four agents, which in turn means that it is still optimal to target the maximum number of groups whenever $r > 1$.

²³With some abuse of notation we omit (1) since we only care about one period of patience and we substitute it by (t) to show that the function is now considered with respect to t .

For an infinitely patient planner we focus on the probability of successful diffusion P_B . By Theorem 2, if $p > 1/2$ it is optimal to target one group, whereas for $p < 1/2$ it is optimal to target the maximum number of groups, which is $m^* = \min\{s/2, t/2\}$. Hence, under the optimal targeting strategies, P_B is equal to:

$$P_B^*(t) = \begin{cases} \frac{r^{\frac{n}{2}} - r^{\frac{s}{2}}}{r^{\frac{n}{2}} - 1} & \text{if } p > 1/2 \\ \frac{r^{\frac{n}{t}} - r^{\frac{n}{t}-1}}{r^{\frac{n}{t}} - 1} & \text{if } p < 1/2 \text{ and } t \leq n/2 \\ \frac{r^{\frac{n}{s}} - r}{r^{\frac{n}{s}} - 1} & \text{if } p < 1/2 \text{ and } t > n/2 \end{cases}$$

where recall that $r = \frac{p}{1-p}$.

The following lemma determines the properties of this function with respect to t . As in most of the analysis, we assume the function to be continuous and twice differentiable in t , and then we evaluate it only on the integer values of t .

Lemma 5. P_B^* has the following properties with respect to t :

- It is increasing in t .
- For $p > 1/2$, it is concave in t
- For $p > 1/2$, it is convex in t if $t \leq n/2$ and concave if $t > n/2$

To explore the meaning of these results we should define explicitly the expected profits of the planner. We use a simple and intuitive expression that captures the main flavor of the results. Of course, one could argue that a more sophisticated profit function might yield better intuitions, but we try to keep the same as notation-free as possible style of the rest of the paper. Therefore, we define the expected profits, as a function of t , as follows:

$$E\Pi(t) = \begin{cases} \pi_i EN_B^*(t) - c(t) & \text{if the planner is impatient} \\ \pi_p P_B^*(t) - c(t) & \text{if the planner is infinitely patient} \end{cases}$$

where the constant π_i is a fixed benefit that the planner receives from each adopter of action B in the first period after targeting and π_p is a fixed benefit that the planner receives if action B captures the whole population. The cost function $c(t)$ is assumed to be linear in t , i.e. $c(t) = kt$, for $k \geq 0$. As usually, $E\Pi(t)$, EN_B^* , P_B^* and c are smooth functions, for which we look only at integer values of t . The following proposition characterizes the optimal number of targeted agents given k .

Proposition 7. There exist lower bounds \underline{k} and upper bounds \bar{k} , whose values depend on p, n, π_i, π_p , such that the optimal number of targeted agents is:

- $t^* = n$, if $k \leq \underline{k}$,
- $t^* \in (0, n)$, if $\underline{k} < k < \bar{k}$ and
- $t^* = 0$, if $k \geq \bar{k}$

Not surprisingly, we find that there exists a lower bound in the unit cost of targeting, k , below which targeting is so cheap that the planner would like to target all the agents and analogously there exists an upper bound above which targeting is so expensive that the expected returns cannot compensate the planner for any positive number of targeted agents.

Interestingly enough, there is also an intermediate range of unit costs k for which the optimal choice of the planner lies in the interior of $[0, n]$. This result is very intuitive if we recall that targeting additional agents increases the probability of successful diffusion, but never ensures it. Therefore, for a given number of initial adopters, by targeting one more the increase in probability may be so little, that the cost for convincing this agent exceeds the expected increase in profits.

2.7.3 Expected Waiting Time before Absorption

We have analyzed the different optimal targeting strategies, given the planner's level of patience. However, in practice, our definition of long-run may have very different characteristics depending on the parameters of the problem. In order to complement our previous analysis, we turn our attention towards the identification of the expected waiting time before the total diffusion of one of the two actions in the whole population.

Our goal is twofold. On the one hand, we want to quantify the meaning of "infinite patience" of a planner, by characterizing the performance of her optimal choices with respect to the expected waiting time before the population converges to a steady state. On the other hand, we explore potential trade-offs between maximizing the probability of successful diffusion and minimizing the expected waiting time before diffusion occurs.

Not surprisingly, we find that diffusion is expected to occur faster as the number of groups increases. This result is very intuitive, because having more groups (of equal sizes) implies that the size of each group is going to be smaller; and when the groups are smaller then fewer failures are sufficient for their disappearance, feature that leads to faster convergence. This means that, for $p < \frac{1}{2}$ the targeting strategy that maximizes the probability of diffusion, it also minimizes the expected time until absorption, therefore there is no trade-off between the two aspects. Moreover, we see that convergence occurs fast even in absolute terms. On the other hand, for $p > \frac{1}{2}$, our optimal targeting strategy is the one with the maximum expected waiting time, compared to all the symmetric configurations, which for some values of p close to $\frac{1}{2}$ may be unrealistically high in absolute terms. Therefore, there is an important trade-off for the planner.

Furthermore, for the optimal initial configurations if $p > \frac{1}{2}$ the expected waiting time is strictly concave in t and attains an interior maximum, whereas if $p < \frac{1}{2}$, then for $t > s$ the expected waiting time increases as t increases. However, it decreases if $p < \frac{1}{2}$ with $t < s$. On the one hand, for $p > \frac{1}{2}$, this happens because extreme differences in the numbers of adopters and non-adopters lead to quicker absorption compared to more equilibrated configurations, assuming the total number of agents constant. On the other hand, for $p < \frac{1}{2}$ with $t > s$ ($t < s$), increasing t (s) is equivalent to a decrease in s (t), therefore a decrease in the number of groups, and a decrease in the number

of groups means that each group adopters will be larger, hence the expected waiting time before absorption will also be longer.

In addition to this, in most of the cases we see that the expected waiting time becomes larger for values of p close to $\frac{1}{2}$ and decreases sometimes up to one or two orders of magnitude as p approaches either 0 or 1. We see also that the larger the difference between the values of s and t , the more extreme the effect in waiting time of changes in the values of p . The problem becomes particularly important when the planner targets a single group of initial adopters, because then the differences may become extreme (for example look at Figure 2.27 in Appendix C). In addition to this, in absolute terms these changes increase significantly as the population increases.

To establish our results on a concrete manner, we use the following result, which comes from Kemeny and Snell (1960). Namely:

Lemma 6 (Kemeny and Snell (1960)). *Consider a random walk with states $\{0, 1, \dots, n\}$, where both barriers 0 and n are absorbing. If the probability of moving to the right (from i to $i + 1$) is p , the probability of moving to the left (from i to $i - 1$) is q and $r = \frac{p}{q}$, then, starting from state i , the expected waiting time before absorption occurs is:*

$$\tau_{(i,n-i)} = \begin{cases} \frac{1}{p-q} \left[n \frac{r^n - r^{n-i}}{r^n - 1} - i \right] & \text{if } p \neq \frac{1}{2} \\ i(n-i) & \text{if } p = \frac{1}{2} \end{cases} \quad (2.2)$$

Recall that we focus on the case where $q = 1 - p$. Now, we are ready to characterize the expected waiting time for the planner's optimal targeting strategies.

Expected waiting times of optimal configurations

For $r > 1$, recall that the optimal choice for the infinitely patient planner is to target a single group of adopters. In this case the process can be summarized as a random walk with $n/2$ steps and two absorbing barriers, starting from $i = t/2$. Hence the expected waiting time is equal to:

$$\tau(t, s | r > 1) = \tau_{(\frac{t}{2}, \frac{s}{2})} = \frac{1}{2p-1} \left[\frac{n}{2} \left(\frac{r^{n/2} - r^{s/2}}{r^{n/2} - 1} \right) - \frac{t}{2} \right] = \frac{1}{2p-1} \left[\frac{s}{2} - \frac{n}{2} \left(\frac{r^{s/2} - 1}{r^{n/2} - 1} \right) \right]$$

The following result summarizes the properties of the above expression:

Remark 1. For $r > 1$, the expected waiting time before absorption, $\tau(t, s | r > 1)$, is

- i. Strictly concave in s and in t ,
- ii. Given p and n , attains interior maximum $t^* = n - s^* = n - 2 \frac{\ln(r^{n/2}-1) - \ln(\ln r) - \ln(n/2)}{\ln r}$
- iii. $\lim_{r \rightarrow 1^+} t^* = \frac{n}{2}$ and $\lim_{r \rightarrow 1^+} \tau(t^*, s^* = \frac{n}{2}) = \frac{n^2}{16}$
- iv. $\lim_{r \rightarrow +\infty} t^* = 0$ and $\lim_{r \rightarrow +\infty} \tau(t^*, s^* = n-2) = \frac{n}{2} - 1$.

◁

Notice, that t^* may not be an even integer number. In this case we have to compare the two closest even integers to t^* , in order to identify the size of t that would give the maximum expected time before absorption occurs. As r diverges to infinity, t^* approaches 0, which is not in the domain of t , hence we check the expected time at $t = 2$.

When p is close to $\frac{1}{2}$, the relation between the expected time of absorption, τ , and the probability of success, p , is a bit more sensitive to the different parameter values. This makes it hard to provide a general result. However, we have analyzed numerically all circular networks with up to one million agents and we have found that, for all the values of p when we can target sufficiently many agents and for values of p sufficiently far away from $1/2$ the expected time of absorption is decreasing in p . Sufficiently far away has different meaning depending on the size of the network. For populations larger than 30 agents is less than 0.6, for populations larger than 100 agents is below 0.55 and for populations larger than 1000 it gets so close to $1/2$ that in practice it is true everywhere. Moreover, this decrease may be as big as one or two orders of magnitude. Figures 2.27 and 2.28 depict the expected time of absorption and its derivative with respect to p respectively, for different values of t , in a population with 200 agents.

For $r < 1$ we must differentiate between the cases where $t < s$ and $t > s$. We concentrate in cases where there is no problem of divisibility, i.e. $s_i = s/m$ and $t_i = t/m$ respectively. For $t < s$, after some simplifications, the expected waiting time is equal to:

$$\tau(t, s | r < 1, t < s) = \tau_{(1, \frac{s}{t})} = \frac{1}{2p-1} \left[\frac{s}{n-s} - \frac{n}{n-s} \left(\frac{r^{\frac{s}{n-s}} - 1}{r^{\frac{n}{n-s}} - 1} \right) \right]$$

Remark 2. For $r < 1$ and $t \leq s$, the expected waiting time, $\tau(t, s | r < 1, t < s)$, is

- i. Strictly decreasing in t and strictly increasing in s ,
- ii. $\lim_{r \rightarrow 0} \tau = 1$ and $\lim_{r \rightarrow 1^-} \tau = \frac{s}{n-s}$,
- iii. $\tau(t = \frac{n}{2}, s = \frac{n}{2}) = \tau_{(1,1)} = 1$ and $\lim_{s \rightarrow n} \tau = -\frac{1}{2p-1}$

◁

The first result is quite intuitive. Increasing $t(s)$, decreases (increases) $\frac{s}{t}$ and therefore decreases (increases) the length of the random walk associated with the process. Decreasing (Increasing) the length of the walk, decreases (increases) the expected time before absorption occurs.

More interesting is part (3) of Remark 2 which shows that the minimum and the maximum expected waiting time do not depend on the size of the population, but only on the probability of success. This result provides a natural upper bound for the expected waiting time, which does not explode as the population grows.

For the relation between τ and p we repeat the same numerical analysis we performed before for $r > 1$ and for all circular networks with up to one million agents. We find that for $s > \frac{n}{2}$ the expected waiting time is strictly increasing in p . For, $s = \frac{n}{2}$ is stable and equal to 1. The behavior is

qualitatively identical irrespective of the size of the population. Figures 2.29 and 2.30 show the typical relation for different values of t .

Notice that, from parts (1) and (2) of Remark 2, we find that the maximum expected time of absorption occurs as p approaches $\frac{1}{2}$ and $t = 2$ and is equal to $\frac{n}{2} - 1$. This is exactly equal to the previous case, described in Remark 1, where $r \rightarrow \infty$ and $t = 2$, which is the fastest among the configurations that maximize the expected waiting time. This shows that a comparison between worst case scenarios for $r < 1$ and $r > 1$, shows that the slowest of all the worst case scenarios when $r < 1$ is still faster than the fastest of the worst case configurations for $r > 1$. This provides a notion of how much faster are the optimal configurations for $r < 1$ with respect to those for $r > 1$.

For $t > s$, the intuitions are analogous to the previous case. The expected waiting time before absorption is equal to:

$$\tau(t, s | r < 1, t > s) = \tau_{(\frac{t}{s}, 1)} = \frac{1}{2p - 1} \left[1 - \frac{n}{s} \frac{r - 1}{r^{\frac{n}{s}} - 1} \right]$$

Remark 3. For $r < 1$ and $t \geq s$, the expected waiting time, $\tau(t, s | r < 1, t > s)$, is:

i. Strictly increasing in t and strictly decreasing in s ,

ii. $\lim_{r \rightarrow 0} \tau = \lim_{r \rightarrow 1^-} \tau = \frac{t}{s}$

iii. $\lim_{s \rightarrow 0} \tau = +\infty$ and $\lim_{s \rightarrow \frac{n}{2}} \tau = 1$

◁

As expected, the first result is the contrary to the first result of Remark 2. The justification is a straightforward reversal of the argument used before. An important issue that arises here is that the expected time is not always increasing as p increases. In fact, it seems to be increasing in the beginning up to some point and then decreasing. The point where it gets maximized varies a lot as we alter p, s and t . Again, a typical behavior can be found in Figures 2.31 and 2.32. The shape of the figures is very similar for all networks we checked (all networks with up to one million agents). An interesting feature is expected waiting is minimized as p approaches 0 and $1/2$ with the two limits being always exactly equal.

Comparison between configurations with different number of groups

Now, we turn our attention to the comparison between configurations consisting of different number of groups. We analyze the simplest of these cases, where all the groups are symmetric. Recall that, when groups are symmetric, then the optimal strategy for an infinitely patient planner is, if $r > 1$, to locate them in a single group and if $r < 1$ to spread them to as many groups as possible. Moreover, even when we allow for asymmetric configurations, the optimal choices again tend to be symmetric. However, it is not clear yet which is the effect of this choice on the expected waiting time before absorption occurs. By Lemma 6, this is equal to:

$$\tau_{(\frac{t}{2m}, \frac{s}{2m})} = \frac{1}{2p-1} \left[\frac{s}{2m} - \frac{n}{2m} \left(\frac{r^{\frac{s}{2m}} - 1}{r^{\frac{n}{2m}} - 1} \right) \right]$$

Proposition 8. *For all $r \neq 1$ and for symmetric configurations, i.e. $s_1 = \dots = s_m = \frac{s}{m}$ and $t_1 = \dots = t_m = \frac{t}{m}$, the expected waiting time is strictly decreasing in the number of groups, m .*

Numerical examples can be found in Figures 2.33 and 2.34, which show that this decrease has an exponential shape. Even though we have restricted our attention to symmetric configurations, this proposition provides an interesting intuition regarding the trade-off between maximizing the probability of successful diffusion and minimizing the expected waiting time before diffusion occurs. On the one hand, for $r > 1$, the choice that maximizes the probability of successful diffusion is the one with the longest expected waiting time. On the other hand, for $r < 1$, the planner's choice that maximizes the probability of successful diffusion is also the one minimizing the expected waiting time until diffusion. Theoretically, these findings are in line with standard results on finite random walks with absorbing barriers, where increasing equally the number of steps in both directions leads to an increase in the expected waiting time before absorption, but it has an ambiguous effect on the probability of absorption in each one of the two barriers.

In our problem, this result stresses out the fact that the strategy of concentrating all the initial adopters in one group when r is high despite maximizing the probability of successful diffusion it may slow down the procedure significantly. This should be taken into account by a planner when trying to quantify the meaning of being infinitely patient. In practical terms, the waiting time may be so long that even a very patient planner would find it unrealistic to wait until diffusion occurs. Nevertheless, when r is low, the optimal choice for the planner is also the one which minimizes the expected waiting time. This facilitates the decision of the planner since there is no trade-off between the two characteristics of the procedure.

Trade-off between probability of success & expected waiting time

One might argue that the trade-off between probability of successful diffusion and expected waiting time before diffusion occurs is not that obvious. The planner might care about the waiting time before the total diffusion of action B , instead of the time before total diffusion of one of the two actions (either A or B). For this reason, we perform a simple exercise where the planner obtains a positive gain from total diffusion of her preferred action and a negative one from total diffusion of the alternative action. We find that for extreme values (either high or low) of p the optimal strategies of the planner might be determined by the speed at which diffusion occurs, rather than by the probability of successful diffusion. In fact, irrespectively of the level of patience, if p is very low the planner prefers to locate all the agents together, whereas if it is very high prefers to spread them as much as possible.

Formally, consider a planner that cares only about total diffusion of one of the action. In particular, the planner receives a positive payoff that is normalized to +1 for each period after total diffusion of action B occurs and a negative payoff that is normalized to -1 for each period after total diffusion

of action A occurs. In addition, assume that the planner discounts future payoffs at a rate $\delta \in (0, 1)$. Hence, her expected payoff will be $\Pi = P_B \sum_{T=\tau}^{+\infty} \delta^T - (1 - P_B) \sum_{T=\tau}^{+\infty} \delta^T = (2P_B - 1) \frac{\delta^\tau}{1-\delta}$, where τ is the expected time before total diffusion of one action occurs and P_B is the probability of successful diffusion. To keep things simple we consider only the case where the planner can target groups of equal size, therefore the payoff depends only on the number of groups. Hence $P_B = \frac{r \frac{n}{2m} - r \frac{s}{2m}}{r \frac{n}{2m} - 1}$ and $\tau = \frac{1}{2p-1} \left[\frac{s}{2m} - \frac{n}{2m} \left(\frac{r \frac{s}{2m} - 1}{r \frac{n}{2m} - 1} \right) \right]$

Let us now try to find the maximum of this function. We know from Proposition 5 that $\frac{dP_B}{dm}$ is positive when $r < 1$ and negative when $r > 1$ and from Proposition 8 that $\frac{d\tau}{dm} < 0$ always. If one differentiates with respect to m will see that for most values of r there is a trade-off between probability and time, but we can easily see that on the one hand $\lim_{r \rightarrow 0} \frac{d\Pi}{dm} < 0$, therefore for any level of patience if the probability of success is sufficiently low, then the planner prefers to concentrate all the agents in one group. On the other hand, $\lim_{r \rightarrow \infty} \frac{d\Pi}{dm} > 0$ which means that, again for any level of patience, if the probability of success is sufficiently high, then the planner prefers to spread the targeted agents to as many groups as possible.

The intuition is similar to that described for the impatient planner, in the sense that high (low) probability of success make the planner optimistic (pessimistic) enough regarding the successful diffusion that he cares about this to happen as quickly (slowly) as possible. This can be also seen as an additional attempt to characterize optimal strategies for intermediate levels of patience, but should be approached with caution since it is only a partial result.

2.7.4 $q \neq 1 - p$: Conformity, Switching Cost and Possibility of Inertia

Until now, we have focused on the case where $q = 1 - p$. In practice this meant that if two neighbors were using different actions at some period, then in the next period either they would both choose action B , which would happen with probability p , or would both use action A , which would happen with probability $1 - p$. This assumption rules out several realistic problems, which we mention here.

Nevertheless, our whole analysis, for both the patient and the impatient planner, is based not on the value of p itself, but on the value of r , which we defined as the relative likelihood of action B being more successful than action A , more formally as $r = \frac{p}{q}$. Despite the fact that, for simplicity reasons, we concentrated on the case where $q = 1 - p$, this need not be the case. Our results remain unchanged if we substitute the condition $p > \frac{1}{2}$ ($p < \frac{1}{2}$ respectively) with $p > q$ ($p < q$ respectively). In both cases, the conditions can be summarized by the value of r and in fact whether $r > 1$ or $r < 1$. In this extension we discuss three realistic scenarios, where relaxing this assumption would be necessary and therefore our results are appropriate.

Possibility of Inertia

With our previous description we have ruled out the possibility of both actions yielding the same payoff. A plausible modification would be to allow both actions to be equally successful in some

states of nature. This would lead to a combination of configurations with transition probabilities as shown in Figure 2.10.

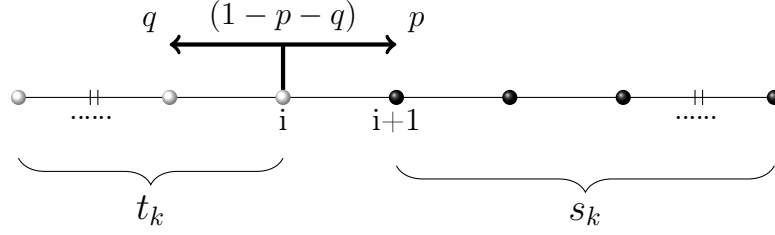


Figure 2.10: Transition probabilities under the possibility of inertia

Focusing on i and $i + 1$, who would be the only agents who could change their decision (in this part of the network), we would obtain the following possible cases:



Figure 2.11: The left figure shows agents' i and $i + 1$ choices at period τ . The other three figures show the choices of the agents at period $\tau + 1$, after (ii) success, (iii) failure, (iv) draw at period τ .

As we have already mentioned, the results are still valid if we define $r = \frac{p}{q}$.

Switching Cost

In the example of technology adoption, we have disregarded the effect of switching costs. Implementing a new technology assumes the purchase of new machinery, or the effort of learning how to use the new technology efficiently. Therefore, a farmer in order to decide to pay this cost must have observed the new technology to have been sufficiently better than the one she already uses. Sufficiently better in this setting can be translated as follows. Agent i who uses action B at period τ changes to action A if the payoff of action A at that period was sufficiently higher, i.e. $\pi_A^\tau > \pi_B^\tau + c$, where π_B^τ is the payoff of B at that period and c the switching cost. Let us call q the probability of this realization. Analogously, agent $i + 1$ who uses action A , changes to action B if $\pi_B^\tau > \pi_A^\tau + c$, which occurs with probability p . Now, if $|\pi_A - \pi_B| < c$ which happens with probability $1 - p - q$,²⁴ then both agents keep using the same action. The alternative action did not seem to be successful enough to convince them to abandon their current technology. This scenario is along the same lines with those we have already mentioned. Hence, our results are also suitable under the presence of switching costs.

Conformity

In the environment we have considered until now, as long as an agent was observing both actions the probability of choosing one over the other was the same, irrespectively of how many of her neighbors

²⁴We assume that the payoffs are defined such that the switching cost is high enough to ensure that $p + q < 1$.

were choosing it. However, our mechanism is also suitable for describing cases where the choices of the agents do not depend only on the performance of the actions, but also on the number of their neighbors (including themselves) who used each action in the previous period. For example, think of a mechanism where there are two cutoff values in the payoff distribution of action B , $b_2 > b_1$, such that in period $\tau + 1$ action B is chosen by those agents who observed

- One or more neighbors choosing action B in period τ , if $b > b_2$,
- Two or more neighbors choosing action B in period τ , if $b_2 > b > b_1$,
- Exactly three neighbors choosing action B in period τ , if $b < b_1$.

where b is the realized payoff of action B in period τ

The three realizations occur with probabilities p_1, p_2, p_3 respectively, where $p_1 + p_2 + p_3 = 1$. The values of the probabilities depend on the payoff distribution of action B . The dynamics of the network can be summarized by Figure 2.12.

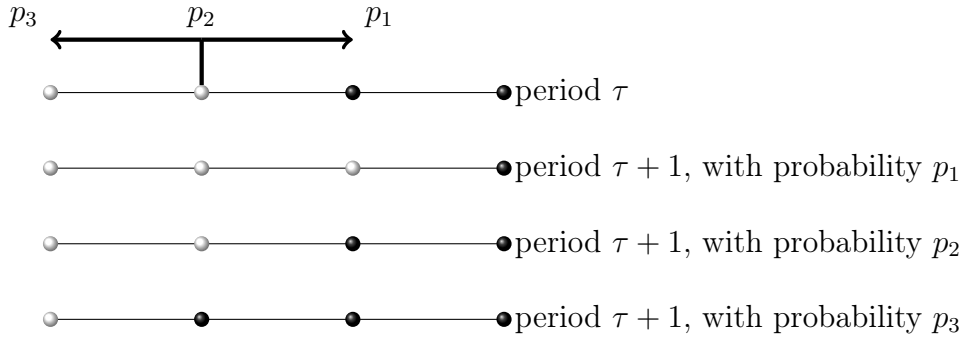


Figure 2.12: Possible configurations and transition probabilities under the presence of conformity.

This mechanism introduces a notion of *conformity* in the network, since the more of ones neighbors choose an action, the more probable is that she chooses it as well. It is apparent that all of our results still hold in this modified environment.

2.7.5 On the Importance of Centrality – The Line

In this section we turn our attention towards the *linear* network. The only difference between the linear and the circle networks is that in the linear network agents 1 and n are not connected between them, so they have only one neighbor. Formally, $N_i = \{i - 1, i, i + 1\}$ for $i = 2, \dots, n - 1$, whereas $N_1 = \{1, 2\}$ and $N_n = \{n - 1, n\}$.

This structure introduces a notion of centrality in the network. Mainly, this is not because some agents have different number of neighbors, but because there is only one path that connects indirectly each two agents. Hence, the agents located closer to the center of the line act as hubs for the transmission of information through the network. This new feature has implications which we are worth discussing.

We only consider the case where the planner can target a single group of agents, of size t , which will be surrounded by two groups of non-adopters with sizes s_1 and s_2 respectively (see Figure 2.13).

For an impatient planner the result is trivial. On the one hand, if $r > 1$ it is optimal to target any segment of the line that does not include any of the corner agents, whereas if $r < 1$ it is optimal to target one of the two corners. This is because for high r the planner expects a success, hence she wants to make the action visible to as many non-adopters as possible. In this case, this number can be at most equal to two, if the planner targets any group that does not include either agent 1 or agent n . To the contrary, when r is low the planner wants to prevent as many adopters as possible from observing the alternative action, therefore she should target one of the two corners. Once again, we will see that this result is in sharp contrast with the the optimal targeting strategy of an infinitely patient planner.

For an infinitely patient planner, we first need to construct $P_B(s_1|s, t, n, r)$ for different values of s_1, s_2 and t . Notice that the only independent variable is s_1 since $s_2 = s - s_1$. Without loss of generality, we consider only the cases where $s_1 \leq s_2$. By symmetry, the remaining cases are completely analogous. We also drop Assumption 1, allowing the groups to have odd number of agents and we discuss the effect this has on the results.

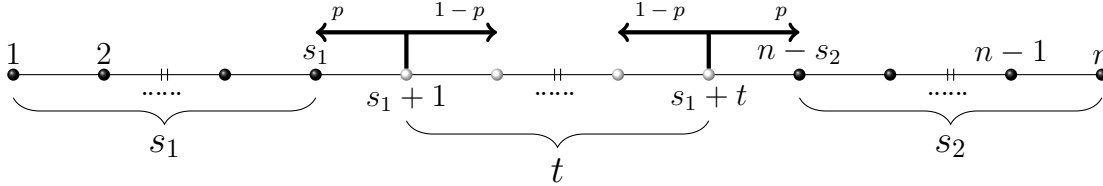


Figure 2.13: A linear network with one group of initial adopters.

The probability of diffusion can be expressed as the product of two random walks with absorbing barriers. The first walk describes the procedure until either the group of type B disappears, or the smaller group of type A disappears. For $r \neq 1$ the first walk is depicted in Figure 2.14. The probability of success in this walk depends on whether t is an odd or an even number. In case the first walk is unsuccessful, action B disappears from the population. In case it is successful, then it is pursued by the random walk in Figure 2.15, which is the same for t being odd or even.

Hence, the probability of diffusion of action B , for $r \neq 1$ is:

$$P_B(s_1|s, t, n, r) = \begin{cases} \frac{r^{(s_1 + \frac{t+1}{2}) - r s_1} r^{n - r(n - t - 2s_1)}}{r^{(s_1 + \frac{t+1}{2}) - 1} r^n - 1} & \text{if } t \text{ odd} \\ \frac{r^{(s_1 + \frac{t}{2}) - r s_1} r^{n - r(n - t - 2s_1)}}{r^{(s_1 + \frac{t}{2}) - 1} r^n - 1} & \text{if } t \text{ even} \end{cases} \quad (2.3)$$

and for $r = 1$:

$$P_B(s_1|s, t, n, r = 1) = \begin{cases} \frac{(t+1)(t+2s_1)}{(t+2s_1+1)n} & \text{if } t \text{ odd} \\ \frac{t}{n} & \text{if } t \text{ even} \end{cases} \quad (2.4)$$

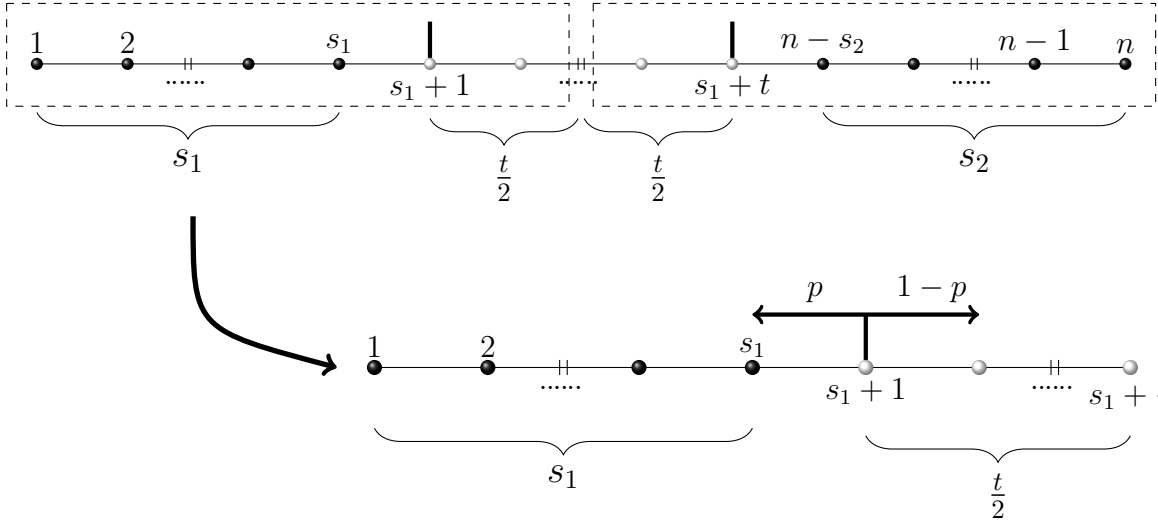


Figure 2.14: The first random walk.

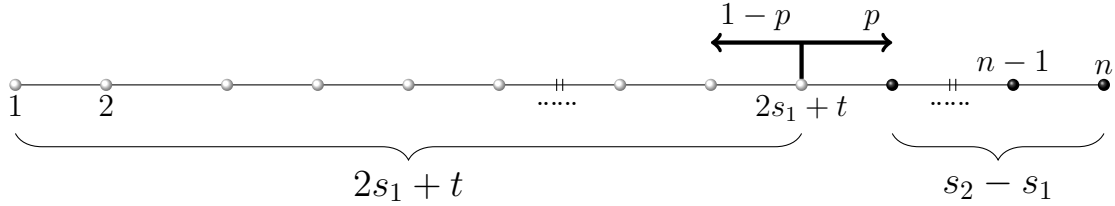


Figure 2.15: The random walk that follows successful absorption in the first walk.

We see that the probability depends slightly on whether t is an odd or an even number. This is because an odd number of initial adopters provides one additional step before the disappearance of action B from the society.

Proposition 9. *If the number t of initial adopters is **even** then:*

- *If $r < 1$ then target the middle, i.e. $s_1 = s_2$ if s even, or $s_1 = s_2 - 1$ if s odd.*
- *If $r > 1$ then target a corner, i.e. $s_1 = 0$.*
- *If $r = 1$ then the probability does not depend on s_1 .*

Proposition 10. *If the number t of initial adopters is **odd** then:*

- *If $r < 1$ then target the middle, i.e. $s_1^* = s_2$ if s even, or $s_1^* = s_2 - 1$ if s odd.*
- *If $r = 1$ then target the middle, i.e. $s_1^* = s_2$ if s even, or $s_1^* = s_2 - 1$ if s odd.*
- *If $r > 1$ then $s_1^* \in \{\lfloor g(r, t) \rfloor, \lceil g(r, t) \rceil\}$,²⁵ for $g(r, t) = \frac{\ln[r^{1/2} + (r-1)^{1/2}]}{\ln r} - \frac{t}{2}$. If $P_B(s_1 = \lfloor g(r, t) \rfloor) > P_B(s_1 = \lceil g(r, t) \rceil)$ then $s_1^* = \lfloor g(r, t) \rfloor$ and vice versa.*

²⁵ $\lfloor g(r, t) \rfloor = \max\{m \in \mathbb{Z} | m \leq g(r, t)\}$ is the floor function of g and $\lceil g(r, t) \rceil = \max\{m \in \mathbb{Z} | m \geq g(r, t)\}$ is the ceiling function of g

These two propositions clarify the difference between having groups with odd or even number of agents. We see that for an odd number of agents, the exposition of the results is slightly more complicated, without providing additional insights. The following corollary provides some more concrete results regarding the cases where t is odd.

Corollary 2. *For t being an **odd** number*

- i. If $t = 1$ it is never optimal to target the corner.*
- ii. For $t \geq 3$ it is optimal to target the corner whenever $r \geq 1.618$.*
- iii. For all $t \geq 3$, there exists $\hat{r} \leq 1.618$ such that for $r > \hat{r}$ it is optimal to target the corner.*

In particular, it seems that the results for odd t are substantially different than for even t only when the number of targeted agents is small. In particular, if we can target only one agent, we never want her to be in the corner of the network. This happens because when we target only one agent, she is never safe for more than one period, meaning that a failure in the first period leads to the disappearance of the action. Moreover, after a positive shock an agent located in the corner can affect the choice of only one additional agent, instead of two. Nevertheless, the problem becomes unimportant when the number of targeted agents is sufficiently large or r is sufficiently high.

2.7.6 On the Importance of Centrality – The Star

The *star* network is a very special case, because there is a unique agent -*the center*- who performs as hub for the information transmission in the network. She observes the actions and outcomes of all the other -*peripheral*- agents, while everyone else observes only her. Formally, $N_1 = \{1, \dots, n\} = N$ while $N_i = \{1, i\}$ for all $i \in \{2, \dots, n\}$. This extreme form of centrality turns out to be crucial in the present setting, making always optimal to target the central agent (independently of the value of r).

Namely, if the central agent (call it “agent 1”) is targeted together with l more peripheral agents, then $P_B(1, l) = \frac{p}{1-p(1-p)}$,²⁶ which is bounded below by p and does not depend on l , as long as $l > 0$. On the other hand, if the central agent is not targeted, then for any number l' of targeted peripheral agents $P_B(not1, l') = pP_B(1, l') = \frac{p^2}{1-p(1-p)}$ which is bounded above by p and again does not depend on l' .

Hence, it is apparent that it is always optimal to target the central agent. Targeting, also, a peripheral agent increases the probability of diffusion, because it secures the action from disappearing in case of a failure in the first period. Targeting more than one peripheral agents does not improve the chances of successful diffusion, since all of them would transmit to the central agent information that she is already aware of.

²⁶ $P_B(1, l) = p + (1-p)P_B(not1, l) = p + (1-p)pP_B(1, l) \Rightarrow P_B(1, l) = \frac{p}{1-p(1-p)}$

2.7.7 On the Importance of Centrality – Other Networks

In this section, we slightly modify the network structure by adding links to the circular network. We run a set of numerical simulations to test the robustness of our results to the addition of a few links and we find that the conclusions remain valid. In particular, if $p > 1/2$, it is always optimal to target only one group of connected agents, with the optimal location of the group depending on the position of the additional links. Conversely, if $p < 1/2$, it is almost always optimal to target the subset of agents that minimizes the number of successful draws needed to capture the whole population.

For the simulation exercise, we consider a network of $n = 20$ agents located around a circle and to this network we add either one or two undirected links. Then we choose the number of targeted agents t and we calculate all possible initial configurations given n and t . We consider only groups of even size, so as to make results more easily comparable with the previous analysis. Subsequently, we simulate the process 4000 times for each initial configuration and for eight different values of p and we estimate P_B for each of the cases. An example of the Matlab code used for the simulations can be found in Appendix D.

Subsequently, we use a standard t-test to check whether the differences between the estimated values are statistically significant. As one would expect, we cannot always find a unique optimal targeting strategy. This happens either because indeed some configurations have very similar probabilities of successful diffusion, or because those probabilities are either very low or very high and therefore we would need a very large sample to choose between the alternatives. In either case, the differences are so small that they are practically unimportant.

In the following tables, we show two particular examples of the simulated processes. For expositional clarity, we report only the configurations that yielded the highest estimated P_B . Nevertheless, all potential optimal strategies of each case have the same characteristics.²⁷

Observe that, for $p = 0.45$ it is optimal to concentrate all the agents together, rather than trying to minimize the number of shocks needed to capture the whole population. The intuition behind this result is that as the network becomes more connected, at least for sufficiently high p (but still lower than $1/2$), the gain from spreading the agents becomes less important compared to the gain from protecting them by concentrating them all together. It is also important to mention that for high values of p , targeting one connected group of agents seems very robust to the modifications of the network structure. Moreover, we can see that agents with a higher number of connections are natural candidates for being targeted only for low values of p . Having noticed that, it seems that an important parameter for the characterization of optimal targeting strategies in more general network structures would be their diameter, which in some cases might be equally important to their degree distribution.

²⁷A complete report of the simulations' results is available upon request.

Table 2.1: Optimal targets in a circle with one additional link, for $n = 20$, $t = 4$.

Added Link	Probability of success (p)							
	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85
1-3	5,6,15,16	1,2,11,12	10,11,20,1	4,5,14,15	7,8,9,10	8,9,10,11	6,7,8,9	3,4,5,6
1-5	1,2,11,12	12,13,20,1	4,5,12,13	10,11,20,1	7,8,9,10	5,6,7,8	11,12,13,14	9,10,11,12
1-7	6,7,14,15	6,7,14,15	12,13,20,1	1,2,13,14	7,8,9,10	4,5,6,7	4,5,6,7	11,12,13,14
1-9	1,2,17,18	2,3,14,15	8,9,12,13	16,17,20,1	1,2,3,4	1,2,3,4	1,2,3,4	2,3,4,5
1-11	8,9,18,19	6,7,16,17	6,7,16,17	8,9,10,11	1,2,19,20	10,11,12,13	15,16,17,18	4,5,6,7

Table 2.2: Optimal targets in a circle with two additional links, for $n = 20$, $t = 8$.

Added Links	Probability of success (p)					
1-5 AND	0.25	0.35	0.45	0.55	0.65	0.75
1-7	1,2,11-14,19,20	10-13,18-20,1	10-13,18-20,1	13-20	8-15	10-17
1-9	2,3,8,9,14,15,20,1	8,9,12,13,16,17,20,1	8-15	8-15	10-17	14-20,1
1-11	8,9,12,13,18,19,20,1	5,6,9,10,15,16,17,18	8,9,12,13,18,19,20,1	15-20,1,2	11-18	10-17
1-13	4,5,10,11,14,15,20,1	6,7,10,11,16,17,20,1	10-15,20,1	7-14	6-13	13-20
1-15	1,2,9-12,19,20	8-11,18,19,20,1	1,2,9-12,19,20	8-15	8-15	14-20,1
1-17	5,6,9,10,15,16,19,20	5,6,9,10,15-18	10-17	10-17	8-15	11-18

2.8 Appendix B - Proofs

Proof of Proposition 5. Under (A1), $\frac{s}{m}$ and $\frac{t}{m}$ are even numbers. Then, the process is equivalent to having a line of $\frac{n}{2m}$ agents, consisting of one group of $\frac{t}{2m}$ adjacent agents choosing B and another group of $\frac{s}{2m}$ adjacent agents choosing A .

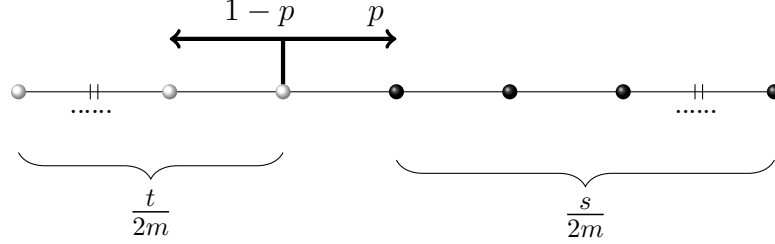


Figure 2.16: The random walk that describes the process in the symmetric case.

By Lemma 4, the probability of successful diffusion becomes:

$$P_B(m|s, t, n, r) = \frac{r^{\frac{n}{2m}} - r^{\frac{s}{2m}}}{r^{\frac{n}{2m}} - 1}$$

Despite the fact, that we are interested only in the integer values of m, t and n , the function $P_B(\cdot)$ is well-defined and smooth for all $r \neq 1$ and $m \geq 1$. Hence, we can check its monotonicity by differentiating with respect to m .

$$\begin{aligned} \frac{dP_B}{dm} &= \frac{\left[r^{\frac{n}{2m}} \ln r \left(-\frac{n}{2m^2} \right) - r^{\frac{s}{2m}} \ln r \left(-\frac{s}{2m^2} \right) \right] (r^{\frac{n}{2m}} - 1) - (r^{\frac{n}{2m}} - r^{\frac{s}{2m}}) \left[r^{\frac{n}{2m}} \ln r \left(-\frac{n}{2m^2} \right) \right]}{(r^{\frac{n}{2m}} - 1)^2} \\ &= \frac{\ln r}{(r^{\frac{n}{2m}} - 1)^2} \left[\left(-\frac{n}{2m^2} r^{\frac{n}{2m}} + \frac{s}{2m^2} r^{\frac{s}{2m}} \right) (r^{\frac{n}{2m}} - 1) - (r^{\frac{n}{2m}} - r^{\frac{s}{2m}}) \left(-\frac{n}{2m^2} r^{\frac{n}{2m}} \right) \right] \\ &= \frac{\ln r}{(r^{\frac{n}{2m}} - 1)^2} \left(-\frac{n}{2m^2} r^{\frac{2n}{2m}} + \frac{s}{2m^2} r^{\frac{s+n}{2m}} + \frac{n}{2m^2} r^{\frac{n}{2m}} - \frac{s}{2m^2} r^{\frac{s}{2m}} + \frac{n}{2m^2} r^{\frac{2n}{2m}} - \frac{n}{2m^2} r^{\frac{s+n}{2m}} \right) \\ &= \frac{\ln r}{2m^2 (r^{\frac{n}{2m}} - 1)^2} \left(sr^{\frac{s+n}{2m}} + nr^{\frac{n}{2m}} - sr^{\frac{s}{2m}} - nr^{\frac{s+n}{2m}} \right) \\ &= \frac{\ln r}{2m^2 (r^{\frac{n}{2m}} - 1)^2} \left[sr^{\frac{s}{2m}} (r^{\frac{n}{2m}} - 1) - nr^{\frac{n}{2m}} (r^{\frac{s}{2m}} - 1) \right] \\ &= \frac{\ln r}{2m^2 (r^{\frac{n}{2m}} - 1)^2} (r^{\frac{n}{2m}} - 1) (r^{\frac{s}{2m}} - 1) \left(\frac{sr^{\frac{s}{2m}}}{r^{\frac{s}{2m}} - 1} - \frac{nr^{\frac{n}{2m}}}{r^{\frac{n}{2m}} - 1} \right) \end{aligned}$$

If we call $\frac{s}{2m} = s'$ and $\frac{n}{2m} = n'$, then the following lemma helps us conclude the argument.

Lemma 7. $f(x) = \frac{2mxr^x}{r^x - 1}$ is strictly increasing for $x \geq 1$, for all $r \neq 1$ and $m \geq 1$

Proof. Let $r \neq 1$ and $m \geq 1$, then

$$\begin{aligned}\frac{df}{dx} &= \frac{2m}{(r^x - 1)^2} [(r^x + xr^x \ln r)(r^x - 1) - xr^x(r^x) \ln r] \\ &= \frac{2m}{(r^x - 1)^2} (r^{2x} - r^x - xr^x \ln r) = \frac{2mr^x}{(r^x - 1)^2} (r^x - 1 - x \ln r) > 0 \text{ for all } x \geq 1\end{aligned}$$

To show this, we define $g(x) = r^x - 1 - x \ln r$, which is strictly increasing for $x \geq 1$ because $\frac{dg}{dx} = r^x \ln r - \ln r = \ln r(r^x - 1) > 0$. So it attains minimum for $x = 1$, which is $g(1) = r - 1 - \ln r$. Moreover, $g(1) > 0$ for all $r \neq 1$ because it holds that $h(r) = r - 1 - \ln r > 0$ for all $r \neq 1$. This holds because $\frac{dh}{dr} = 1 - \frac{1}{r}$ is strictly positive when $r > 1$ and strictly negative when $r < 1$. So, h attains global minimum for $r = 1$, with value $h(1) = 0$. Hence, $g(x) > 0$ for all $x \geq 1$, which means that also $\frac{df}{dx} > 0$ for all $x \geq 1$ and this concludes the argument. \square

By Lemma 7, given that $n > s$, we get that $(\frac{sr^{\frac{s}{2m}}}{r^{\frac{s}{2m}} - 1} - \frac{nr^{\frac{n}{2m}}}{r^{\frac{n}{2m}} - 1}) < 0$ always, so we can conclude that $\frac{dP_B}{dm} < 0$ if $r > 1$ and $\frac{dP_B}{dm} > 0$ if $r < 1$. Hence, for $r > 1$ the $P_B(m|s, t, n, r)$ is decreasing in m , so $\arg \max_m P_B(m|s, t, n, r) = 1$, i.e. the optimal choice is to target a single group of initial adopters. On the other hand, for $r < 1$, P is increasing in m , so we would like to split the initial adopters in as many groups as possible, i.e. $\arg \max_m P_B(m|s, t, n, r) = \min\{s/2, t/2\}$. \square

Proof of Proposition 6. First, we have to construct the probability of successful diffusion. For $r \neq 1$, the process again can be described as a sequence of random walks with absorbing barriers. At the beginning, we have a random walk of $(s_1 + t_1)/2$ nodes, starting from node $t_1/2$, until it disappears either t_1 or s_1 . By Lemma 4, the probability of successful absorption of this walk is $\frac{r^{\frac{s_1+t_1}{2}} - r^{\frac{s_1}{2}}}{r^{\frac{s_1+t_1}{2}} - 1}$. In case of successful absorption we get a random walk of $n/2$ nodes starting from the node $(t + 2s_1)/2$. Otherwise, in case of unsuccessful absorption we get a random walk of $n/2$ nodes as well, but starting from node $(t - 2t_1)/2$. Again by Lemma 4, the probabilities of successful absorption in these two scenarios are $\frac{r^{\frac{n}{2}} - r^{\frac{n-t-2s_1}{2}}}{r^{\frac{n}{2}} - 1}$ and $\frac{r^{\frac{n}{2}} - r^{\frac{s+2t_1}{2}}}{r^{\frac{n}{2}} - 1}$ respectively. If the second walk, in either of the two scenarios, is unsuccessful then action B disappears. Figure 2.17 depicts the process we just described. Notice that (A1) solves all the problems of divisibility.

The histories that lead to full diffusion of action B are (i) success in both the first and the second walk and (ii) failure in the first and success in the second walk. Therefore the probability of successful diffusion for $r \neq 1$ can be written as follows:

$$P_B(s_1, t_1|s, t, n, r) = \frac{r^{\frac{s_1+t_1}{2}} - r^{\frac{s_1}{2}}}{r^{\frac{s_1+t_1}{2}} - 1} \frac{r^{\frac{n}{2}} - r^{\frac{n-t-2s_1}{2}}}{r^{\frac{n}{2}} - 1} + \frac{r^{\frac{s_1}{2}} - 1}{r^{\frac{s_1+t_1}{2}} - 1} \frac{r^{\frac{n}{2}} - r^{\frac{s+2t_1}{2}}}{r^{\frac{n}{2}} - 1}$$

Now, we compute the derivatives with respect to t_1 and s_1 . As usually, we are only interested in integer points, but the function P_B is a well-behaved smooth function for $r \neq 1$, so we can study its monotonicity.

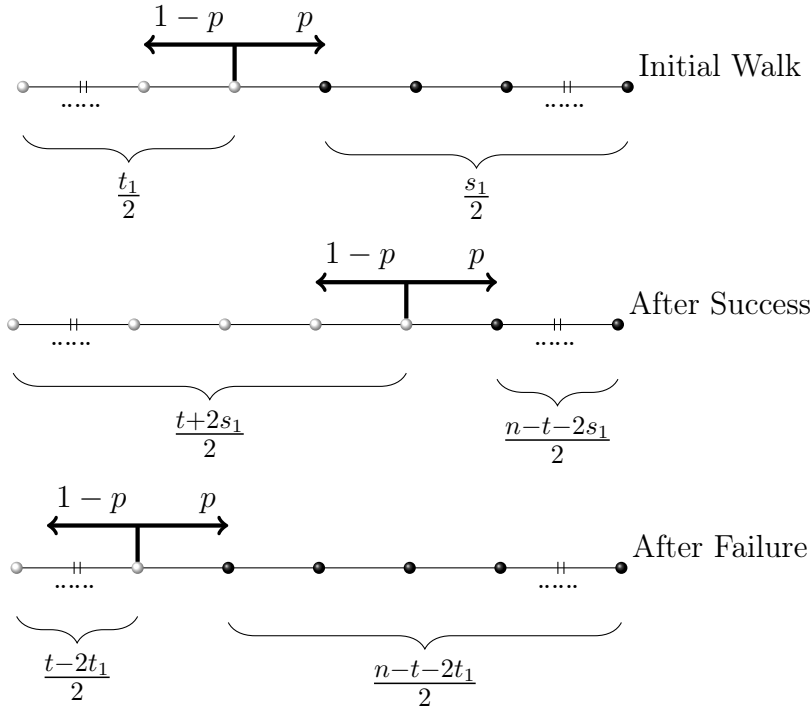


Figure 2.17: The random walks that describe the process in the asymmetric case with two groups.

$$\begin{aligned}
\frac{\partial P_B}{\partial s_1} &= \left[\frac{r^{\frac{n}{2}} \left(r^{\frac{t_1}{2}} - 1 \right)}{r^{\frac{n}{2}} - 1} \right] \left[\frac{\left(\frac{\ln r}{2} r^{\frac{s_1}{2}} + \frac{\ln r}{2} r^{\frac{-s_1-t}{2}} \right) \left(r^{\frac{s_1+t_1}{2}} - 1 \right) - \left(r^{\frac{s_1}{2}} - r^{\frac{-s_1-t}{2}} \right) \frac{\ln r}{2} r^{\frac{s_1+t_1}{2}}}{\left(r^{\frac{s_1+t_1}{2}} - 1 \right)^2} \right] + \\
&+ \left(\frac{r^{\frac{n}{2}} - r^{\frac{s+2t_1}{2}}}{r^{\frac{n}{2}} - 1} \right) \left[\frac{\frac{\ln r}{2} r^{\frac{s_1}{2}} \left(r^{\frac{s_1+t_1}{2}} - 1 \right) - \left(r^{\frac{s_1}{2}} - 1 \right) \frac{\ln r}{2} r^{\frac{s_1+t_1}{2}}}{\left(r^{\frac{s_1+t_1}{2}} - 1 \right)^2} \right] = \\
&= \frac{r^{\frac{n}{2}} \left(r^{\frac{t_1}{2}} - 1 \right) \ln r}{2 \left(r^{\frac{n}{2}} - 1 \right) \left(r^{\frac{s_1+t_1}{2}} - 1 \right)^2} \left(r^{\frac{2s_1+t_1}{2}} + r^{\frac{-t+t_1}{2}} - r^{\frac{s_1}{2}} - r^{\frac{-t-s_1}{2}} - r^{\frac{2s_1+t_1}{2}} + r^{\frac{-t+t_1}{2}} \right) + \\
&+ \frac{\left(r^{\frac{n}{2}} - r^{\frac{s+2t_1}{2}} \right) \ln r}{2 \left(r^{\frac{n}{2}} - 1 \right) \left(r^{\frac{s_1+t_1}{2}} - 1 \right)^2} \left(r^{\frac{2s_1+t_1}{2}} - r^{\frac{s_1}{2}} - r^{\frac{2s_1+t_1}{2}} + r^{\frac{s_1+t_1}{2}} \right) = \\
&= \frac{\ln r r^{\frac{n}{2}} \left(r^{\frac{t_1}{2}} - 1 \right) \left[2r^{\frac{-t+t_1}{2}} - r^{\frac{s_1}{2}} - r^{\frac{-t-s_1}{2}} + r^{\frac{s_1}{2}} - r^{\frac{-t+2t_1+s_1}{2}} \right]}{2 \left(r^{\frac{n}{2}} - 1 \right) \left(r^{\frac{s_1+t_1}{2}} - 1 \right)^2} = \\
&= \frac{\ln r \left(r^{\frac{t_1}{2}} - 1 \right) \left(2r^{\frac{n-t+t_1}{2}} - r^{\frac{n-t-s_1}{2}} - r^{\frac{s+2t_1}{2}} \right)}{2 \left(r^{\frac{n}{2}} - 1 \right) \left(r^{\frac{s_1+t_1}{2}} - 1 \right)^2} = \frac{\ln r \left(r^{\frac{t_1}{2}} - 1 \right) r^{\frac{s-s_1}{2}} \left(2r^{\frac{s_1+t_1}{2}} - 1 - r^{s_1+t_1} \right)}{2 \left(r^{\frac{n}{2}} - 1 \right) \left(r^{\frac{s_1+t_1}{2}} - 1 \right)^2} = \\
&= \frac{\ln r \left(r^{\frac{t_1}{2}} - 1 \right) r^{\frac{s-s_1}{2}} \left[- \left(r^{\frac{s_1+t_1}{2}} - 1 \right)^2 \right]}{2 \left(r^{\frac{n}{2}} - 1 \right) \left(r^{\frac{s_1+t_1}{2}} - 1 \right)^2} = - \frac{\ln r \left(r^{\frac{t_1}{2}} - 1 \right) r^{\frac{s-s_1}{2}}}{2 \left(r^{\frac{n}{2}} - 1 \right)} \begin{cases} < 0 & \text{if } r > 1 \\ > 0 & \text{if } r < 1 \end{cases}
\end{aligned}$$

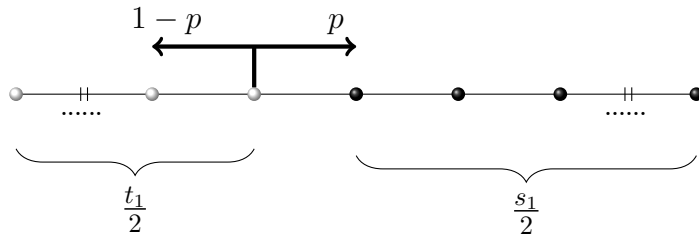
Analogously for t_1 we have:

$$\begin{aligned}
\frac{\partial P_B}{\partial t_1} &= \frac{\ln r \left[(r^{\frac{n+s_1}{2}} - r^{\frac{s-s_1}{2}}) r^{\frac{t_1}{2}} (r^{\frac{s_1+t_1}{2}} - 1) - (r^{\frac{n+s_1}{2}} - r^{\frac{s-s_1}{2}}) (r^{\frac{t_1}{2}} - 1) r^{\frac{s_1+t_1}{2}} \right]}{2(r^{\frac{n}{2}} - 1)(r^{\frac{s_1+t_1}{2}} - 1)^2} + \\
&+ \frac{\ln r \left[(r^{\frac{s_1}{2}} - 1)(-2r^{\frac{s+2t_1}{2}})(r^{\frac{s_1+t_1}{2}} - 1) - (r^{\frac{s_1}{2}} - 1)(r^{\frac{n}{2}} - r^{\frac{s+2t_1}{2}}) r^{\frac{s_1+t_1}{2}} \right]}{2(r^{\frac{n}{2}} - 1)(r^{\frac{s_1+t_1}{2}} - 1)^2} = \\
&= \frac{\ln r}{2(r^{\frac{n}{2}} - 1)(r^{\frac{s_1+t_1}{2}} - 1)^2} \left(r^{\frac{n+s_1}{2}} - r^{\frac{s-s_1}{2}} \right) \left(-r^{\frac{t_1}{2}} + r^{\frac{s_1+t_1}{2}} \right) + \\
&+ \frac{\ln r}{2(r^{\frac{n}{2}} - 1)(r^{\frac{s_1+t_1}{2}} - 1)^2} (r^{s_1/2} - 1) \left(-r^{\frac{s+s_1+3t_1}{2}} + 2r^{\frac{s+2t_1}{2}} - r^{\frac{n+s_1+t_1}{2}} \right) = \\
&= \frac{\ln r \left(r^{\frac{s_1}{2}} - 1 \right) \left(-r^{\frac{s+t_1-s_1}{2}} - r^{\frac{s+s_1+3t_1}{2}} + 2r^{\frac{s+2t_1}{2}} \right)}{2(r^{\frac{n}{2}} - 1) \left(r^{\frac{s_1+t_1}{2}} - 1 \right)^2} = \\
&= \frac{\ln r \left(r^{\frac{s_1}{2}} - 1 \right) r^{\frac{s+t_1}{2}}}{2r^{\frac{s_1}{2}} (r^{\frac{n}{2}} - 1) \left(r^{\frac{s_1+t_1}{2}} - 1 \right)^2} \left(2r^{\frac{s_1+t_1}{2}} - r^{\frac{2s_1+2t_1}{2}} - 1 \right) = \\
&= \frac{\ln r \left(r^{\frac{s_1}{2}} - 1 \right) r^{\frac{s+t_1}{2}} \left[- \left(r^{\frac{s_1+t_1}{2}} - 1 \right)^2 \right]}{2r^{\frac{s_1}{2}} (r^{\frac{n}{2}} - 1) \left(r^{\frac{s_1+t_1}{2}} - 1 \right)^2} = - \frac{\ln r \left(r^{\frac{s_1}{2}} - 1 \right) r^{\frac{s+t_1}{2}}}{2r^{\frac{s_1}{2}} (r^{\frac{n}{2}} - 1)} \begin{cases} < 0 & \text{if } r > 1 \\ > 0 & \text{if } r < 1 \end{cases}
\end{aligned}$$

Hence, given that $0 \leq s_1 \leq s_2$ and $0 \leq t_1 \leq t_2$ we conclude that for $r > 1$ the optimal targeting strategy is $(s_1, t_1) = (0, 0)$, whereas for $r < 1$ it is $s_2 - s_1 \leq 2$ and $t_2 - t_1 \leq 2$.

□

Proof of Theorem 2. For the case of $r > 1$ we proceed by induction. First, we recall the result by Proposition 5, which states that if we can target up to two groups, then the optimal choice is to concentrate all the initial adopters in one group. Remember also that $s_1 \leq s_2 \leq s_3$ and $t_1 \leq t_2 \leq t_3$. Now suppose that we can target up to three groups ($m \leq 3$). Then again at first we are interested in the two smallest groups of each type and we have the following random walk:



The system fluctuates in this direction until either s_1 or t_1 disappears. Depending on the successful or unsuccessful absorption of this process we get one of the configurations depicted in Figure 2.18, with only two groups of each type left.

By Proposition 5, in both of these cases we know that the optimal choice would be to eliminate one of the two groups of initial adopters. Hence, we would like to choose s_2 and t_2 (as functions of s_1 and t_1 respectively), in such a way that the probability of diffusion is maximized in both of these cases. Recalling that $s_1 \leq s_2 \leq s_3$ and $t_1 \leq t_2 \leq t_3$, we see that this can be achieved if $s_2 = s_1$ and $t_2 = t_1$, where the optimal s_1 and t_1 remained to be determined. Notice that, by construction, $s_3 = s - s_1 - s_2$ and $t_3 = t - t_1 - t_2$.

So now, we can rewrite the probability of diffusion as a function of s_1 and t_1 only.

$$P_B(s_1, t_1 | s, t, n, r, m = 3) = \frac{r^{\frac{t_1+s_1}{2}} - r^{\frac{s_1}{2}} r^{\frac{n}{2}-r} r^{\frac{n-t-3s_1}{2}}}{r^{\frac{t_1+s_1}{2}} - 1} \frac{r^{\frac{n}{2}} - r^{\frac{s+3t_1}{2}}}{r^{\frac{n}{2}} - 1} + \frac{r^{\frac{s_1}{2}} - 1}{r^{\frac{t_1+s_1}{2}} - 1} \frac{r^{\frac{n}{2}} - r^{\frac{s+3t_1}{2}}}{r^{\frac{n}{2}} - 1}$$

Like before we study the monotonicity of the function with respect to s_1 and t_1

$$\begin{aligned} \frac{\partial P_B}{\partial s_1} &= \frac{\ln r}{2(r^{\frac{n}{2}} - 1)(r^{\frac{s_1+t_1}{2}} - 1)^2} \left[\left(r^{\frac{t_1}{2}} - 1 \right) \left(r^{\frac{n+s_1}{2}} + 2r^{\frac{s-2s_1}{2}} \right) + \left(r^{\frac{n}{2}} - r^{\frac{s+3t_1}{2}} \right) r^{\frac{s_1}{2}} \right] \left(r^{\frac{s_1+t_1}{2}} - 1 \right) - \\ &\quad - \frac{\ln r}{2(r^{\frac{n}{2}} - 1)(r^{\frac{s_1+t_1}{2}} - 1)^2} \left[\left(r^{\frac{t_1}{2}} - 1 \right) \left(r^{\frac{n+s_1}{2}} - r^{\frac{s-2s_1}{2}} \right) + \left(r^{\frac{n}{2}} - r^{\frac{s+3t_1}{2}} \right) \left(r^{\frac{s_1}{2}} - 1 \right) \right] r^{\frac{s_1+t_1}{2}} = \\ &= \frac{\ln r}{2(r^{\frac{n}{2}} - 1)(r^{\frac{s_1+t_1}{2}} - 1)^2} \left(r^{\frac{t_1}{2}} - 1 \right) \left[\left(r^{\frac{n+s_1}{2}} + 2r^{\frac{s-2s_1}{2}} \right) \left(r^{\frac{s_1+t_1}{2}} - 1 \right) - \left(r^{\frac{n+s_1}{2}} - r^{\frac{s-2s_1}{2}} \right) r^{\frac{s_1+t_1}{2}} \right] + \\ &\quad + \frac{\ln r}{2(r^{\frac{n}{2}} - 1)(r^{\frac{s_1+t_1}{2}} - 1)^2} \left(r^{\frac{n}{2}} - r^{\frac{s+3t_1}{2}} \right) \left[r^{\frac{s_1}{2}} \left(r^{\frac{s_1+t_1}{2}} - 1 \right) - \left(r^{\frac{s_1}{2}} - 1 \right) r^{\frac{s_1+t_1}{2}} \right] = \\ &= \frac{\ln r}{2(r^{\frac{n}{2}} - 1)(r^{\frac{s_1+t_1}{2}} - 1)^2} \left(r^{\frac{t_1}{2}} - 1 \right) \left(3r^{\frac{s+t_1-s_1}{2}} - 2r^{\frac{s-2s_1}{2}} - r^{\frac{s+s_1+3t_1}{2}} \right) = \\ &= \frac{r^{\frac{s}{2}} \ln r}{2(r^{\frac{n}{2}} - 1)(r^{\frac{s_1+t_1}{2}} - 1)^2} \left(r^{\frac{t_1}{2}} - 1 \right) \left(3r^{\frac{t_1-s_1}{2}} - 2r^{\frac{-2s_1}{2}} - r^{\frac{s_1+3t_1}{2}} \right) = \\ &= \frac{r^{\frac{s}{2}} \ln r}{2(r^{\frac{n}{2}} - 1)(r^{\frac{s_1+t_1}{2}} - 1)^2} \left(r^{\frac{t_1}{2}} - 1 \right) \left[-\frac{\left(r^{\frac{s_1+t_1}{2}} - 1 \right)^2 \left(r^{\frac{s_1+t_1}{2}} + 2 \right)}{r^{s_1}} \right] = \\ &= -\frac{r^{\frac{s}{2}} \ln r \left(r^{\frac{t_1}{2}} - 1 \right) \left(r^{\frac{s_1+t_1}{2}} + 2 \right)}{2(r^{\frac{n}{2}} - 1)r^{s_1}} < 0 \text{ for } r > 1 \end{aligned}$$

Hence the optimal choice is $s_1 = s_2 = 0$ and $s_3 = s$.

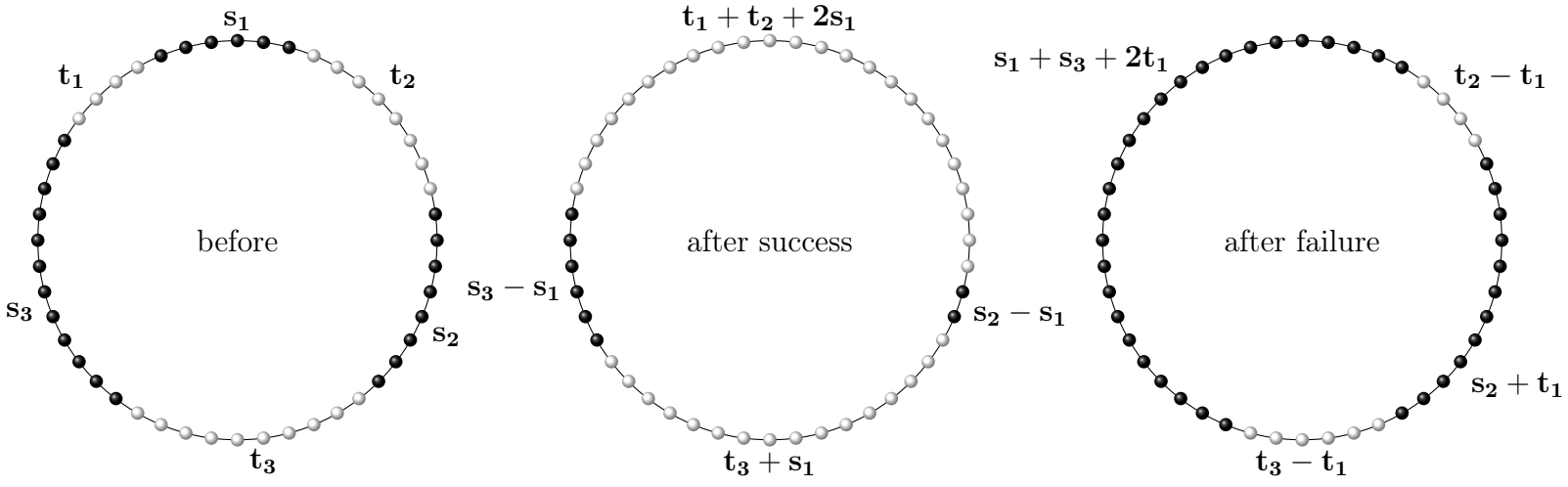


Figure 2.18: Configurations with 3 groups of adopters.

Analogously for t_1 we get the following:

$$\begin{aligned}
\frac{\partial P_B}{\partial t_1} &= \frac{\ln r}{2 \left(r^{\frac{n}{2}} - 1\right) \left(r^{\frac{s_1+t_1}{2}} - 1\right)^2} \left(r^{\frac{n+s_1}{2}} - r^{\frac{s-2s_1}{2}}\right) \left[r^{\frac{t_1}{2}} \left(r^{\frac{s_1+t_1}{2}} - 1\right) - r^{\frac{s_1+t_1}{2}} \left(r^{\frac{t_1}{2}} - 1\right)\right] + \\
&+ \frac{\ln r}{2 \left(r^{\frac{n}{2}} - 1\right) \left(r^{\frac{s_1+t_1}{2}} - 1\right)^2} \left(r^{\frac{s_1}{2}} - 1\right) \left[-3r^{\frac{s+3t_1}{2}} \left(r^{\frac{s_1+t_1}{2}} - 1\right) - \left(r^{\frac{n}{2}} - r^{\frac{s+3t_1}{2}}\right) r^{\frac{s_1+t_1}{2}}\right] = \\
&= \ln r \frac{\left(r^{\frac{n+s_1}{2}} - r^{\frac{s-2s_1}{2}}\right) \left(r^{\frac{s_1+t_1}{2}} - r^{\frac{t_1}{2}}\right) + \left(r^{\frac{s_1}{2}} - 1\right) \left(3r^{\frac{s+3t_1}{2}} - 2r^{\frac{s+s_1+4t_1}{2}} - r^{\frac{n+s_1+t_1}{2}}\right)}{2 \left(r^{\frac{n}{2}} - 1\right) \left(r^{\frac{s_1+t_1}{2}} - 1\right)^2} = \\
&= \frac{\ln r \left(r^{\frac{s_1}{2}} - 1\right)}{2 \left(r^{\frac{n}{2}} - 1\right) \left(r^{\frac{s_1+t_1}{2}} - 1\right)^2} \left(3r^{\frac{s+3t_1}{2}} - 2r^{\frac{s+s_1+4t_1}{2}} - r^{\frac{s+t_1-2s_1}{2}}\right) = \\
&= \frac{\ln r \left(r^{\frac{s_1}{2}} - 1\right) r^{\frac{s}{2}} r^{\frac{t_1}{2}}}{2 \left(r^{\frac{n}{2}} - 1\right) \left(r^{\frac{s_1+t_1}{2}} - 1\right)^2} \left(3r^{\frac{2t_1}{2}} - 2r^{\frac{s_1+3t_1}{2}} - r^{\frac{-2s_1}{2}}\right) = \\
&= \frac{\ln r \left(r^{\frac{s_1}{2}} - 1\right) r^{\frac{s}{2}} r^{\frac{t_1}{2}}}{2r^{s_1} \left(r^{\frac{n}{2}} - 1\right) \left(r^{\frac{s_1+t_1}{2}} - 1\right)^2} \left(3r^{\frac{2s_1+2t_1}{2}} - 2r^{\frac{3s_1+3t_1}{2}} - 1\right) = \\
&= \frac{\ln r \left(r^{\frac{s_1}{2}} - 1\right) r^{\frac{s_2-s_1}{2}} r^{\frac{t_1}{2}} \left[-\left(r^{\frac{s_1+t_1}{2}} - 1\right)^2 \left(2r^{\frac{s_1+t_1}{2}} + 1\right)\right]}{2 \left(r^{\frac{n}{2}} - 1\right) \left(r^{\frac{s_1+t_1}{2}} - 1\right)^2} < 0 \text{ for } r > 1.
\end{aligned}$$

The last step comes from the observation that $-2x^3 + 3x^2 - 1 = -(x-1)^2(2x+1)$, where in this case $x = r^{\frac{s_1+t_1}{2}}$. Hence, P_B is always decreasing in t_1 , and given that $t_1 = t_2$ the optimal choice is $t_1 = t_2 = 0$ and $t_3 = t$. This concludes the argument for the case where $m = 3$. We will generalize

this argument by induction.

Formally, given that the argument holds for $m = 3$, it suffices to show that if it holds for $m = k - 1 \geq 3$ then it holds as well for $m = k$.

At the beginning of the process we care only about the two smallest groups of each type s_1 and t_1 and the system fluctuates as in the previous cases until one of the two disappears. Figure 2.19 shows the possible configurations after the disappearance of either s_1 or t_1 . The location of the groups around the network comes without loss of generality.

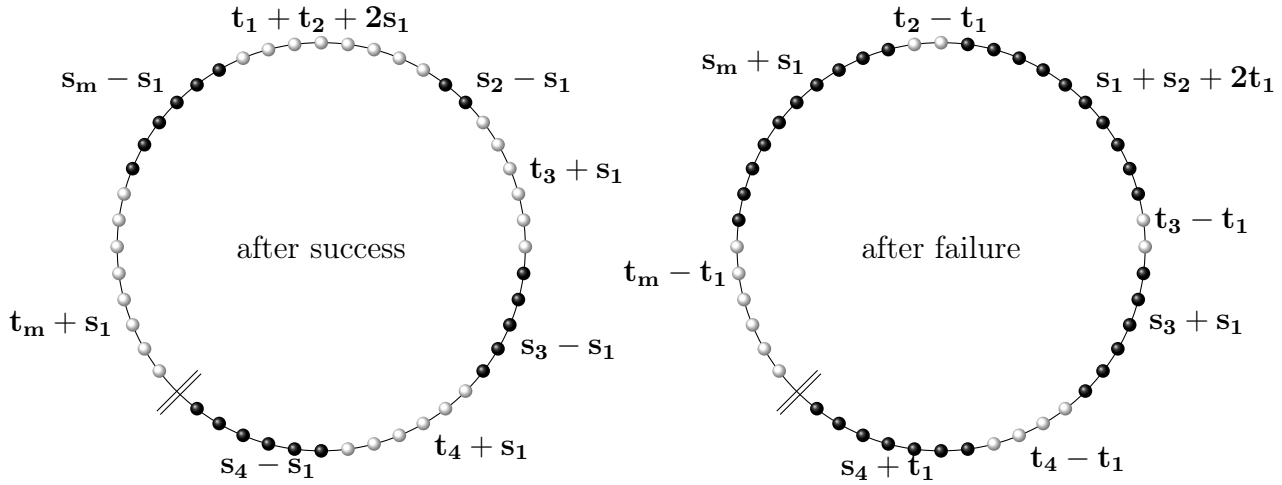


Figure 2.19: Configurations after the disappearance of s_1 or t_1 with m groups.

Given that the argument holds for $k - 1$ groups then we know that $s_1 = \dots = s_{k-1}$ and $t_1 = \dots = t_{k-1}$. Therefore, we only need to find the optimal s_1 and t_1 . The probability of diffusion becomes:

$$P_B(s_1, t_1 | s, t, n, r, m = k) = \frac{r^{\frac{t_1+s_1}{2}} - r^{\frac{s_1}{2}} r^{\frac{n}{2}-r} r^{\frac{n-t-k s_1}{2}}}{r^{\frac{t_1+s_1}{2}} - 1} \frac{r^{\frac{n}{2}} - 1}{r^{\frac{n}{2}} - 1} + \frac{r^{\frac{s_1}{2}} - 1}{r^{\frac{t_1+s_1}{2}} - 1} \frac{r^{\frac{n}{2}} - r^{\frac{s+kt_1}{2}}}{r^{\frac{n}{2}} - 1}$$

By some calculations which are omitted because they are identical to the case where $m = 3$, we get:

$$\frac{\partial P_B}{\partial s_1} = \frac{r^{\frac{s}{2}} \left(r^{\frac{t_1}{2}} - 1 \right) \ln r}{2r^{(k-1)\frac{s_1}{2}} \left(r^{\frac{n}{2}} - 1 \right) \left(r^{\frac{s_1+t_1}{2}} - 1 \right)^2} \left[kr^{\frac{s_1+t_1}{2}} - r^{k\frac{s_1+t_1}{2}} - (k-1) \right] \leq 0 \text{ for } r > 1$$

and equality holds only if $s_1 = t_1 = 0$. For the argument to hold we need $kr^{\frac{s_1+t_1}{2}} - r^{k\frac{s_1+t_1}{2}} - (k-1)$ to be negative. So, let $x = r^{\frac{s_1+t_1}{2}}$ and take the function $f(x) = kx - x^k - (k-1)$ for some $k \geq 3$ and $x \geq 0$. Now, $\frac{df}{dx} = k - kx^{k-1}$ is positive if $x < 1$ and negative if $x > 1$, hence f attains global max at $x = 1$ equal to $f(1) = k - 1^k - (k-1) = 0$, hence $f(x) < 0$ for all $x \neq 1$. Now given that

$x = r^{\frac{s_1+t_1}{2}}$, with $r > 1$ and $s_1, t_1 \geq 0$ the function is always strictly negative and becomes equal to zero only when $s_1 = t_1 = 0$. So, the optimal choice is $s_1 = \dots = s_{k-1} = 0$ and $s_k = s$.

Analogously for t_1 we get that:

$$\frac{\partial P_B}{\partial t_1} = \frac{r^{\frac{s}{2}} r^{\frac{t_1}{2}} \left(r^{\frac{s_1}{2}} - 1 \right) \ln r}{2r^{(k-1)\frac{s_1}{2}} \left(r^{\frac{n}{2}} - 1 \right) \left(r^{\frac{s_1+t_1}{2}} - 1 \right)^2} \left[kr^{(k-1)\frac{s_1+t_1}{2}} - (k-1)r^{k\frac{s_1+t_1}{2}} - 1 \right] \leq 0 \text{ for } r > 1$$

again equality holds only when $s_1 = t_1 = 0$ and to ensure the result we need that $kr^{(k-1)\frac{s_1+t_1}{2}} - (k-1)r^{k\frac{s_1+t_1}{2}} - 1 \leq 0$ for all s_1 and t_1 with equality holding only in case they are both equal to zero. As before, let $x = r^{\frac{s_1+t_1}{2}}$ and define the function $g(x) = kx^{k-1} - (k-1)x^k - 1$. Then $\frac{dg}{dx} = k(k-1)x^{k-2} - k(k-1)x^{k-1}$ which is strictly negative for $x > 1$ and strictly positive for $x < 1$, then g attains unique maximum at $x = 1$ equal to $g(1) = 0$. So $g(x) < 0$ for all $x \neq 1$. Given again that $x = r^{\frac{s_1+t_1}{2}}$ then $x = 1$ only if $s_1 = t_1 = 0$. So again the optimal choices are $t_1 = \dots = t_{k-1} = 0$ and $t_k = t$, which completes the inductive argument. Hence, when $r > 1$, for any possible number of groups m , the optimal choice is to concentrate all the initial adopters in one group, i.e. $s_1 = \dots = s_{m-1} = 0$ and $s_m = s$, as well as $t_1 = \dots = t_{m-1} = 0$ and $t_m = t$.

Now, we turn our attention towards the case where $r < 1$. We tackle this case in a different way. Namely, we construct an upper bound for the probability of successful convergence and we show that the actual probability is equal to this upper bound for the same configurations that this upper bound is maximized. Hence this has to be the maximum value of the probability as well.

In order to proceed, we need to construct the upper bound for the value of the probability of successful diffusion of action B . We solve it first for $t < s$ and then for $s < t$.

Let $t < s$, then allowing for the existence of $m = \frac{t}{2}$ groups, the network will have the form of Figure 2.20. Notice that, the fact that s_i can have size equal to zero, allows us to construct any possible configuration. For example, if $s_1 = 0$ then the two groups next to s_1 merge to one group with four agents. According to this structure, the network will initially follow a random walk with $\frac{s_1}{2}$ black steps and one white. By Lemma 4, the probability of success in this first walk is equal to $\frac{r^{\frac{s_1}{2}+1} - r^{\frac{s_1}{2}}}{r^{\frac{s_1}{2}+1} - 1}$. In Figure 2.20 we also see how the network will look like if the first walk is successful. Unsuccessful absorption in the first walk leads to the disappearance of action B from the network, because all t 's have the same size. After success, the network will move according to the random walk of Figure 2.21, with the probability of success in this walk is equal to $\frac{r^{\frac{s_2}{2}+1} - r^{\frac{s_2-s_1}{2}}}{r^{\frac{s_2}{2}+1} - 1}$.

We depict as well the two possible configurations that arise after success or failure in the second walk (see Figure 2.22). It is important to notice that the probability of successful diffusion after two successes is obviously weakly lower than one and it is strictly lower than one, as long as $s_m - s_1 > 2$, where s_m is the size of the largest group and s_1 is the size of the smallest one.

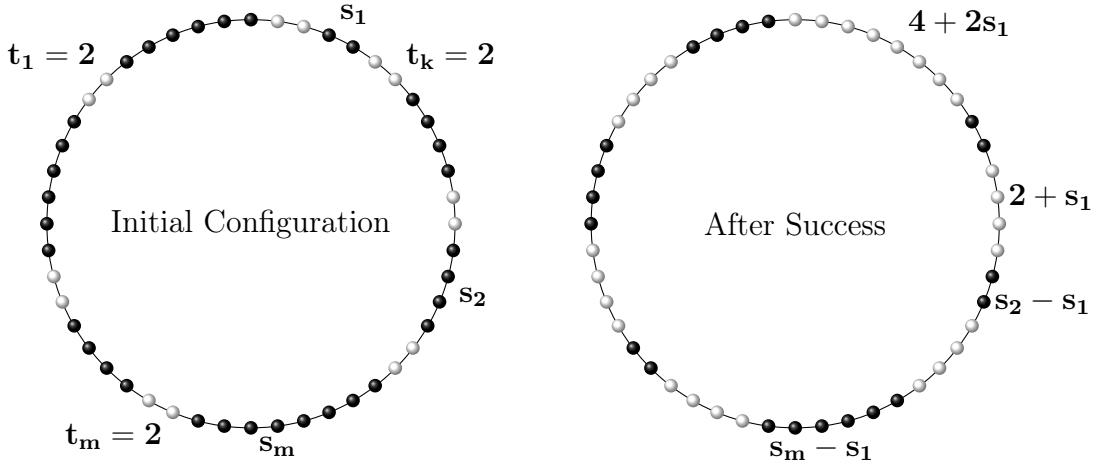


Figure 2.20: General Initial Configuration and Result After Success, for $t < s$.

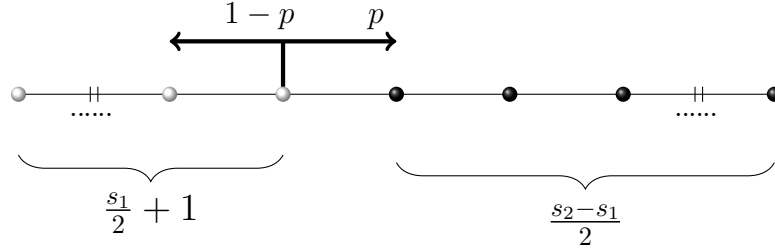


Figure 2.21: Random Walk After Success.

Hence, we can construct the probability of successful diffusion, which is equal to:

$$P_B(\cdot) = \frac{r^{\frac{s_1}{2}+1} - r^{\frac{s_1}{2}}}{r^{\frac{s_1}{2}+1} - 1} \left[\frac{r^{\frac{s_2}{2}+1} - r^{\frac{s_2-s_1}{2}}}{r^{\frac{s_2}{2}+1} - 1} P_B(\cdot|s, s) + \frac{r^{\frac{s_2-s_1}{2}} - 1}{r^{\frac{s_2}{2}+1} - 1} \frac{r^{\frac{n}{2}} - r^{\frac{n-s_1}{2}-1}}{r^{\frac{n}{2}} - 1} \right]$$

where $P_B(\cdot|s, s)$ stands for the probability of diffusion of B after two successes in the first two random walks. Given that $P_B(\cdot|s, s) \leq 1$ we get the following upper bound of P_B , denoted by $\widetilde{P_B}(\cdot)$, which is equal to:

$$\widetilde{P_B}(\cdot) = \frac{r^{\frac{s_1}{2}+1} - r^{\frac{s_1}{2}}}{r^{\frac{s_1}{2}+1} - 1} \left[\frac{r^{\frac{s_2}{2}+1} - r^{\frac{s_2-s_1}{2}}}{r^{\frac{s_2}{2}+1} - 1} + \frac{r^{\frac{s_2-s_1}{2}} - 1}{r^{\frac{s_2}{2}+1} - 1} \frac{r^{\frac{n}{2}} - r^{\frac{n-s_1}{2}-1}}{r^{\frac{n}{2}} - 1} \right]$$

Before performing any calculations it is important to simplify the expression of $\widetilde{P_B}(\cdot)$. Specifically,

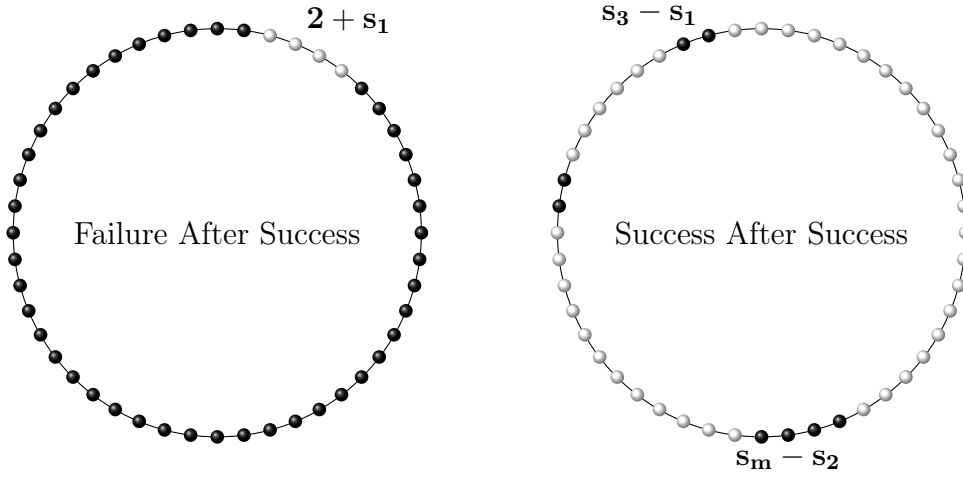


Figure 2.22: Resulting Configurations given Failure or Success in the second random walk, given successful first walk, for $t < s$

$$\begin{aligned}
\widetilde{P_B}(\cdot) &= \frac{r^{\frac{s_1}{2}+1} - r^{\frac{s_1}{2}}}{r^{\frac{s_1}{2}+1} - 1} \left[\frac{r^{\frac{s_2}{2}+1} - r^{\frac{s_2-s_1}{2}}}{r^{\frac{s_2}{2}+1} - 1} + \frac{r^{\frac{s_2-s_1}{2}} - 1}{r^{\frac{s_2}{2}+1} - 1} \frac{r^{\frac{n}{2}} - r^{\frac{n-s_1}{2}-1}}{r^{\frac{n}{2}} - 1} \right] = \\
&= \frac{r-1}{r^{\frac{n}{2}} - 1} \left[\frac{\left(r^{\frac{s_1+s_2}{2}+1} - r^{\frac{s_2}{2}} \right) \left(r^{\frac{n}{2}} - 1 \right) + \left(r^{\frac{n}{2}} - r^{\frac{n-s_1}{2}-1} \right) \left(r^{\frac{s_2}{2}} - r^{\frac{s_1}{2}} \right)}{\left(r^{\frac{s_1}{2}+1} - 1 \right) \left(r^{\frac{s_2}{2}+1} - 1 \right)} \right] = \\
&= \frac{r-1}{r^{\frac{n}{2}} - 1} \left[\frac{\left(r^{\frac{n+s_2}{2}} + r^{\frac{n+s_2-s_1}{2}-1} - r^{\frac{s_2}{2}} - r^{\frac{n}{2}-1} \right) \left(r^{\frac{s_1}{2}+1} - 1 \right)}{\left(r^{\frac{s_1}{2}+1} - 1 \right) \left(r^{\frac{s_2}{2}+1} - 1 \right)} \right] = \\
&= \frac{r-1}{r^{\frac{n}{2}} - 1} \left[\frac{r^{\frac{n}{2}-1} \left(r^{\frac{s_2}{2}+1} - 1 \right)}{r^{\frac{s_2}{2}+1} - 1} + \frac{r^{\frac{n+s_2-s_1}{2}-1} - r^{\frac{s_2}{2}}}{r^{\frac{s_2}{2}+1} - 1} \right] = \\
&= \frac{r-1}{r^{\frac{n}{2}} - 1} \left[r^{\frac{n}{2}-1} + \frac{r^{\frac{n+s_2-s_1}{2}-1} - r^{\frac{s_2}{2}}}{r^{\frac{s_2}{2}+1} - 1} \right]
\end{aligned}$$

Notice that $s_2 = s - s_1 - s_3 - \dots - s_m$, hence $\frac{\partial s_2}{\partial s_1} = -1$. And now we can differentiate $\widetilde{P_B}(\cdot)$ with respect to s_1 .

$$\begin{aligned}
\frac{\partial \widetilde{P_B}(\cdot)}{\partial s_1} &= \frac{r-1}{r^{\frac{n}{2}} - 1} \left[\frac{\left(-r^{\frac{n+s_2-s_1}{2}-1} \ln r + \frac{\ln r}{2} r^{\frac{s_2}{2}} \right) \left(r^{\frac{s_2}{2}+1} - 1 \right) - \left(r^{\frac{n+s_2-s_1}{2}-1} - r^{\frac{s_2}{2}} \right) \left(-\frac{\ln r}{2} r^{\frac{s_2}{2}+1} \right)}{\left(r^{\frac{s_2}{2}+1} - 1 \right)^2} \right] = \\
&= \frac{\ln r (r-1) r^{\frac{s_2}{2}}}{2 \left(r^{\frac{n}{2}} - 1 \right) \left(r^{\frac{s_2}{2}+1} - 1 \right)^2} \left(2r^{\frac{n-s_1}{2}-1} - r^{\frac{n+s_2-s_1}{2}} - 1 \right) > 0, \text{ for } r < 1.
\end{aligned}$$

The fact that the term $2r^{\frac{n-s_1}{2}-1} - r^{\frac{n+s_2-s_1}{2}} - 1$ is always negative is not obvious and is proven here. Substituting s_2 , we can rewrite it as:

$$r^{-s_1-1} \left[2r^{\frac{n+s_1}{2}} - r^{\frac{n+s-s_3-\dots-s_m}{2}+1} - r^{2\frac{s_1}{2}+1} \right]$$

If we denote $x = r^{\frac{s_1}{2}}$ then we get a polynomial of degree two with respect to x . The discriminant of this polynomial is equal to:

$$\Delta = 4r^{2\frac{n}{2}} - 4r^{\frac{n}{2} + \frac{s-s_3-\dots-s_m}{2}+2} = 4r^{\frac{n}{2}} \left(r^{\frac{n}{2}} - r^{\frac{s-s_3-\dots-s_m}{2}+2} \right) = 4r^{\frac{n}{2}} \left(r^{\frac{n}{2}} - r^{\frac{s_1+s_2}{2}+2} \right) < 0$$

Because $r < 1$ and $\frac{n}{2} > \frac{s_1+s_2}{2} + 2$ for $m \geq 3$. For $m = 2$ this holds with equality, but we have already analyzed this case. So, this polynomial has no roots and given that the factor of the quadratic term is negative ($-r$), we can conclude that for $r < 1$ the polynomial is always negative. Therefore, $\widetilde{P_B(\cdot)}$ takes its maximum value when s_1 is maximized. For this value of s_1 , the real probability of successful diffusion is equal to this upper bound as long as $s_m - s_1 \leq 2$. Therefore, remembering that $\widetilde{P_B(\cdot)} \geq P_B(\cdot)$ always, it has to be that $P_B(\cdot)$ is also maximized for when both s_1 is maximized and $s_m - s_1 \leq 2$.

In case m divides s exactly, then the maximum of s_1 is equal to $\frac{s}{m}$ and the optimal choice is $s_1 = \dots = s_m = \frac{s}{m}$. If m does not divide s exactly, then we have $s = mq + d$, where q is the quotient of the division and d is the remainder. In this case, P_B is maximized if we have $m - \frac{d}{2}$ groups of size $q = \frac{s-d}{m}$ and $\frac{d}{2}$ groups with size $q + 2 = \frac{s-d}{m} + 2$, so again the difference in the size of any two groups is no larger than four. We still remain to describe what is the optimal position of the groups that have the two additional agents. The result will become apparent after we analyze the case for $t > s$.

Now, we prove the result for $t > s$ in a completely analogous way. In this case, the initial configuration is as in the left part of Figure 2.23. A success in the first random walk leads to the diffusion of action B , while a failure leads to a configuration as in the right part of the same figure. The probability of success in the first walk is $\frac{r^{\frac{t_1}{2}+1}-r}{r^{\frac{t_1}{2}+1}-1}$. Figure 2.24 shows the possible configurations after successful or unsuccessful absorption in the second random walk, given unsuccessful absorption in the first one. The probability of success in the second walk is $\frac{r^{\frac{t_2}{2}+1}-r^{\frac{t_1}{2}+1}}{r^{\frac{t_2}{2}+1}-1}$. Therefore, we can construct again an upper bound for the probability of successful diffusion, equal to:

$$\widetilde{P_B(\cdot)} = \frac{r^{\frac{t_1}{2}+1}-r}{r^{\frac{t_1}{2}+1}-1} + \frac{r-1}{r^{\frac{t_1}{2}+1}-1} \left[\frac{r^{\frac{t_2}{2}+1}-r^{\frac{t_1}{2}}}{r^{\frac{t_2}{2}+1}-1} \frac{r^{\frac{n}{2}}-r^{\frac{t_1}{2}-1}}{r^{\frac{n}{2}}-1} + \frac{r^{\frac{t_1}{2}}-1}{r^{\frac{t_2}{2}+1}-1} \right]$$

This expression can be transformed in a similar manner as before:

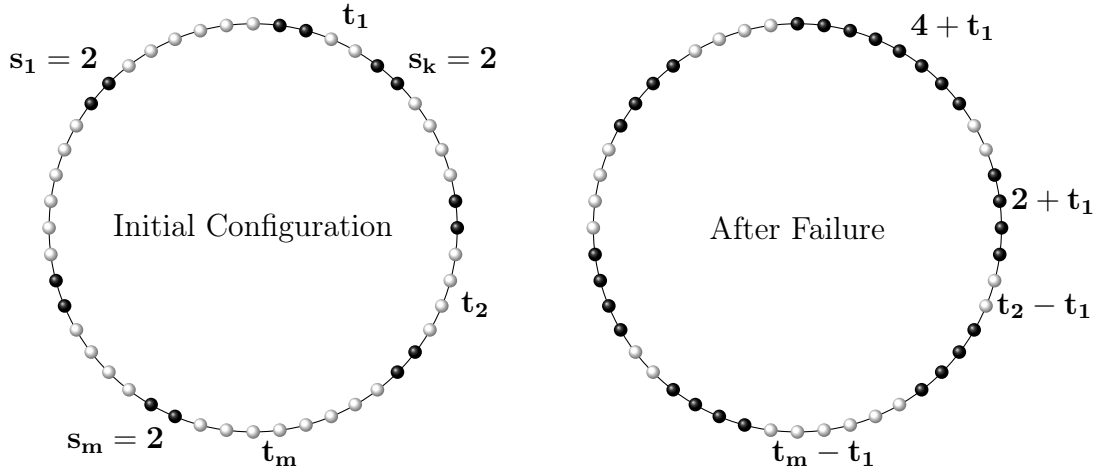


Figure 2.23: General Initial Configuration and Result After Success, for $t > s$.

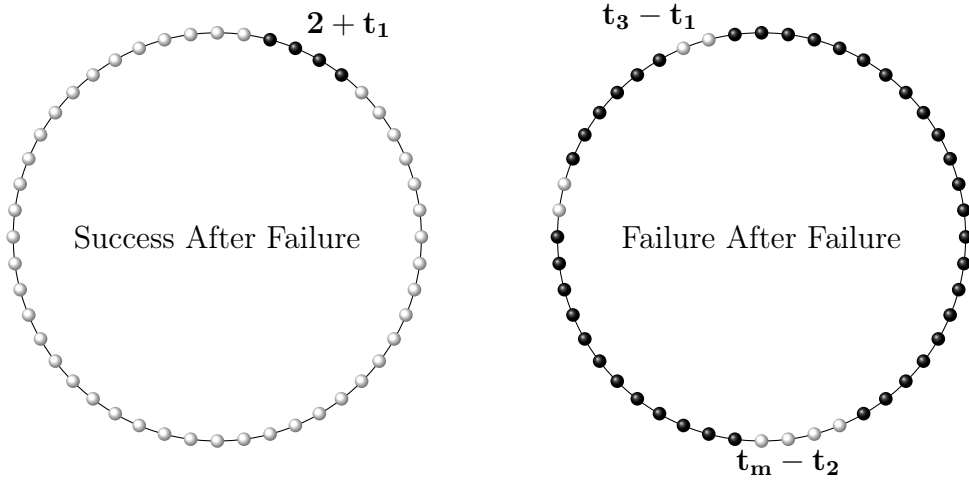


Figure 2.24: Resulting configurations after success or failure in the second random walk, given failure in the first random walk, for $t > s$

$$\begin{aligned}
 \widetilde{P_B(\cdot)} &= \frac{r^{\frac{t_1}{2}+1} - r}{r^{\frac{t_1}{2}+1} - 1} + \frac{r - 1}{r^{\frac{t_1}{2}+1} - 1} \left[\frac{r^{\frac{t_2}{2}+1} - r^{\frac{t_1}{2}}}{r^{\frac{t_2}{2}+1} - 1} \frac{r^{\frac{n}{2}} - r^{\frac{t_1}{2}-1}}{r^{\frac{n}{2}} - 1} + \frac{r^{\frac{t_1}{2}} - 1}{r^{\frac{t_2}{2}+1} - 1} \right] = \\
 &= \frac{1}{r^{n/2} - 1} \left[\frac{(r^{\frac{t_2}{2}+1} - 1)(r^{n/2} - 1) - (r^{\frac{t_2}{2}+1} - r^{\frac{t_1}{2}+1})(r - 1)}{r^{\frac{t_2}{2}+1} - 1} \right] = \\
 &= \frac{1}{r^{n/2} - 1} \left[r^{n/2} - r + (r - 1) \frac{r^{\frac{t_1}{2}+1} - 1}{r^{\frac{t_2}{2}+1} - 1} \right]
 \end{aligned}$$

Notice again, that the upper bound becomes equal to the actual probability if $t_m - t_1 \leq 2$, where t_m is the size of the largest group of type B and t_1 the smallest one.

Now, we can differentiate the expression with respect to t_1 , remembering that $t_2 = t - t_1 - t_3 - \dots - t_m$:

$$\frac{\partial \widetilde{P_B(\cdot)}}{\partial t_1} = \frac{(r-1) \ln r}{2(r^{n/2} - 1)} \left[\frac{r^{\frac{t_1}{2}+1}(r^{\frac{t_2}{2}+1} - 1) + (r^{\frac{t_1}{2}+1} - 1)r^{\frac{t_2}{2}+1}}{(r^{\frac{t_2}{2}+1} - 1)^2} \right] > 0, \text{ for all } r < 1.$$

The upper bound is increasing in t_1 . For this value of t_1 , the real probability of successful diffusion is equal to this upper bound as long as $t_m - t_1 \leq 2$. Therefore, remembering that $\widetilde{P_B(\cdot)} \geq P_B(\cdot)$ always, it has to be that $P_B(\cdot)$ is also maximized for when both t_1 is maximized and $t_m - t_1 \leq 2$.

In case m divides s exactly, then the maximum of t_1 is equal to $\frac{t}{m}$ and the optimal choice is $t_1 = \dots = t_m = \frac{t}{m}$. If m does not divide t exactly, then we have $t = mq + d$, where q is the quotient of the division and d is the remainder. In this case, P_B is maximized if we have $m - \frac{d}{2}$ groups of size $q = \frac{t-d}{m}$ and $\frac{d}{2}$ groups with size $q + 2 = \frac{t-d}{m} + 2$, so again the difference in the size of any two groups is no larger than four.

To complete the proof we need to explain the optimal location of the groups which have the two additional agents. For the case where $t < s$ we need to notice that after successful absorption in the first random walk, now the network consists of $\frac{d}{2}$ groups of each type, where all the groups of type A have exactly two agents. Hence, we fall into the analysis of the case where $t > s$, where we would like the groups of type B to be as equal as possible. In order to succeed this we should have located the groups of type A with more agents as symmetrically as possible around the network.

An example can be illustrated in Figure 2.25. We have targeted 14 out of 48 agents, having seven groups of two agents of type B , three groups of six agents and four groups of four agents of type A . After successful absorption in the first walk, there will be left only three groups of two agents of type A , which we want to be located as symmetrically as possible. For this reason we do not put two groups of six agents one next to the other in the initial configuration. However, notice that we cannot make the configuration arising after success totally symmetric, due to the restriction on the sizes of the groups. But again we want it to be as symmetric as possible, by maximizing the smallest group and minimizing its difference with the largest one. The argument for the case where $t > s$ is completely analogous.

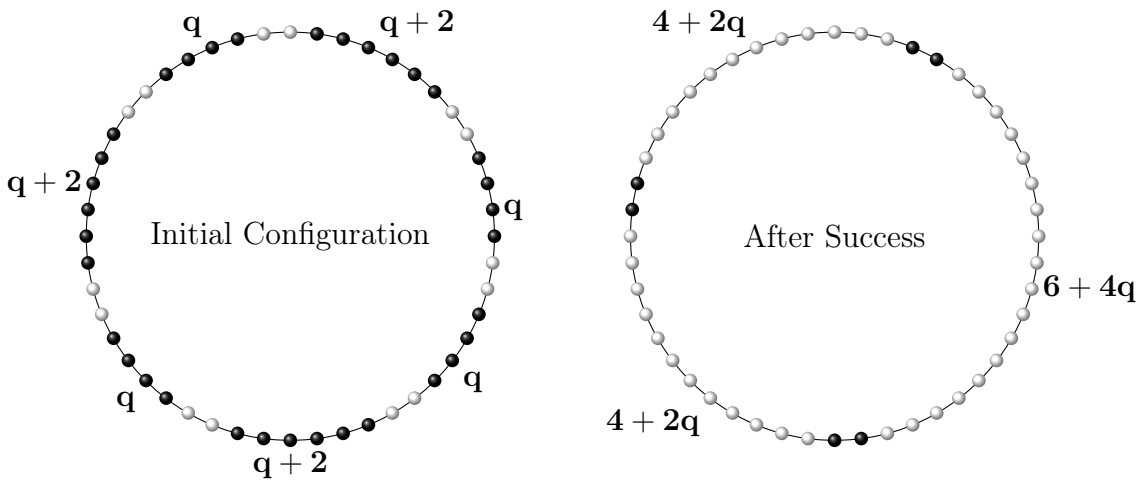


Figure 2.25: Optimal Initial Configuration with $s = mq + d$, for $s > t$.

□

Proof of Lemma 5. We will prove the results for each branch of the function P_B^* separately. First, consider the case of $p > 1/2$, where $P_B^*(t) = \frac{r^{\frac{n}{2}-r^{\frac{s}{2}}}}{r^{\frac{n}{2}-1}}$. One can easily calculate $\frac{dP_B^*}{dt} = \frac{r^{\frac{s}{2}} \ln r}{2(r^{\frac{n}{2}-1})} > 0$ and $\frac{d^2P_B^*}{dt^2} = -\frac{r^{\frac{s}{2}}(\ln r)^2}{4(r^{\frac{n}{2}-1})} < 0$. Now, we turn our attention to the case of $p < 1/2$. If $t \leq n/2$, then $P_B^*(t) = \frac{r^{\frac{n}{t}-r^{\frac{n}{t}-1}}}{r^{\frac{n}{t}-1}}$ from which we get that $\frac{dP_B^*}{dt} = \frac{n(r-1)r^{\frac{n}{t}} \ln r}{t^2(r^{\frac{n}{t}-1})^2} > 0$ and $\frac{d^2P_B^*}{dt^2} = -\frac{n(r-1)r^{\frac{n}{t}} \ln r}{t^3(r^{\frac{n}{t}-1})^3} [2(r^{\frac{n}{t}} - 1) - \frac{n}{t} \ln r (r^{\frac{n}{t}} + 1)] > 0$. The result of the second derivative needs a brief explanation. In fact, we show that the expression inside the brackets $[2(r^{\frac{n}{t}} - 1) - \frac{n}{t} \ln r (r^{\frac{n}{t}} + 1)]$ is positive whenever $r < 1$ (equivalent to $p < 1/2$) irrespectively of n and t . Let us call it $f(r)$ and calculate $f'(r) = \frac{n}{tr} (r^{\frac{n}{t}} - \frac{n}{t} r^{\frac{n}{t}-1} - 1) < 0$ for all $r < 1$. This holds because $g(r) = (r^{\frac{n}{t}} - \frac{n}{t} r^{\frac{n}{t}-1} - 1) < 0$ since $g'(r) = -(\frac{n}{t})^2 r^{\frac{n}{t}-2} \ln r < 0$ for $r < 1$ and $g(1) = -1 < 0$. The fact that f is decreasing and $f(1) = 0$ means $f > 0$ for all $r < 1$, hence the second derivative is positive. Analogously, for $t > n/2$, $P_B^* = \frac{r^{\frac{n}{s}-r}}{r^{\frac{n}{s}-1}}$ from which we get $\frac{dP_B^*}{dt} = \frac{n(r-1)r^{\frac{n}{s}} \ln r}{s^2(r^{\frac{n}{s}-1})^2} > 0$ and $\frac{d^2P_B^*}{dt^2} = \frac{n(r-1)r^{\frac{n}{s}} \ln r}{s^3(r^{\frac{n}{s}-1})^3} [2(r^{\frac{n}{s}} - 1) - \frac{n}{s} \ln r (r^{\frac{n}{s}} + 1)] < 0$, with this last result following directly from the analysis of the previous case.

Therefore, we can conclude that P_B is increasing in t for all values of p , it is concave in t for $p > 1/2$ and also for $p < 1/2$ but in the latter case only when $t \leq n/2$ and finally it is convex in t for $p < 1/2$ when $t > n/2$.

□

Proof of Proposition 7. Although the result is qualitatively the same, we need to proceed separately for the impatient and the infinitely patient planner, as well as for different values of p and t .

For an *impatient* planner, if $p > 1/2$ then the optimal number of targeted agents is

- if $k \leq 2(1-p)\pi_i$, then $t^* = n$,
- if $2(1-p)\pi_i < k \leq 2p\pi_i$, then $t^* = n/2$,
- if $k > 2p\pi_i$, then $t^* = 0$.

If $p > 1/2$ then for $t \leq n/2$ the profit function is $E\Pi = 2p\pi_i t - kt$ which is increasing whenever $2p\pi_i \geq k$ and decreasing otherwise. For $t > n/2$, $E\Pi = n(2p-1)\pi_i + 2(1-p)p\pi_i t - kt$, which is increasing if $2\pi_i(1-p) \geq t$. Notice that, given that $p > 1/2$ it holds that $p > 1-p$. Therefore we can conclude that for $k \leq 2(1-p) < 2p$ the expected profits are always increasing, hence the optimal number of targeted agents, $t^* = n$. Whereas, if $2(1-p) < k < 2p$ the expected profits are increasing until $t = n/2$ and decreasing afterwards, hence $t^* = n/2$. Finally, if $2(1-p) < 2p \leq k$ then the expected profits are always decreasing in t , hence $t^* = 0$.

On the other hand, if $p < 1/2$, $E\Pi = \pi_i t + 2\pi_i(2p-1) - kt$, which simplifies things since it follows immediately that:

- if $k \leq \pi_i$, then $t^* = n$,
- if $k > \pi_i$, then $t^* = 0$.

We now turn attention to the *infinitely patient* planner, for which if $p > 1/2$ then the optimal number of targeted agents is:

- if $k \leq \frac{\pi_p \ln r}{r^{\frac{n}{2}-1}}$, then $t^* = n$,
- if $\frac{\pi_p \ln r}{r^{\frac{n}{2}-1}} < k \leq \frac{\pi_p r^{\frac{n}{2}} \ln r}{r^{\frac{n}{2}-1}}$, then $t^* = P_B^*(\frac{k}{\pi_p})$,²⁸
- if $k > \frac{\pi_p r^{\frac{n}{2}} \ln r}{r^{\frac{n}{2}-1}}$, then $t^* = 0$.

where $P_B^*(\frac{k}{\pi_p})$ is the inverse function of the derivative of P_B^* evaluated at k/π_p .

By Theorem 2, the optimal targeting strategy of the planner is to concentrate all initial adopters in one group. Therefore the expected profits' function has the form $E\Pi(t) = \pi_p \frac{r^{\frac{n}{2}-t} - r^{\frac{n-t}{2}}}{r^{\frac{n}{2}-1}} - kt$ from which it can be calculated $\frac{d^2 E\Pi}{dt^2} = -\frac{\pi_p (\ln r)^2}{4(r^{\frac{n}{2}-1})} r^{\frac{n-t}{2}} - k < 0$. Therefore, expected profit is a strictly concave function of t .

Now it is sufficient to see when the derivative $\frac{dE\Pi}{dt} = \frac{\pi_p \ln r}{2(r^{\frac{n}{2}-1})} r^{\frac{n-t}{2}} - k$ has a root. It is apparent that the derivative has a root if and only if $\frac{dE\Pi}{dt}|_{t=0} < 0$ and $\frac{dE\Pi}{dt}|_{t=n} > 0$, which are satisfied when $\frac{\pi_p \ln r}{2(r^{\frac{n}{2}-1})} < k < \frac{\pi_p \ln r}{2(r^{\frac{n}{2}-1})} r^{\frac{n}{2}}$. In this case the optimal number is the root of this expression, which can be written in general as $t^* = P_B^*(\frac{k}{\pi_p})$, as long as this is an integer. If it is not then the maximum is described in footnote 28. If k is larger or equal than the upper bound then the derivative is always negative, and the optimal solution is $t = 0$. If k is lower or equal than the lower bound then the derivative is always positive and the optimal solution is $t = n$.

Turning our attention to the case of $p < 1/2$, the result is as follows:

- if $k = 0$, then $t^* = n$,
- if $0 < k < \frac{4\pi_p r^2 \ln r}{n(r-1)(r+1)^2}$, then $t^* = P_B^*(\frac{k}{\pi_p})$ with $n/2 \leq t^* < n$
- if $k \geq \frac{4\pi_p r^2 \ln r}{n(r-1)(r+1)^2}$, then $t^* = 0$.

where again $P_B^*(\frac{k}{\pi_p})$ is the inverse function of the derivative of P_B^* evaluated at k/π_p .²⁹ Recall from Lemma 5 that P_B^* is convex until $t = n/2$ and concave thereafter. Hence, P_B^* is first increasing until $t = n/2$ and then decreasing, and attains a maximum at this point. Given that $\frac{dE\Pi}{dt} = \pi_p P_B^* - k$, this derivative is always negative if and only if $\pi_p P_B^*(\frac{n}{2}) \leq k$, or equivalently if $k \geq \frac{4\pi_p r^2 \ln r}{n(r-1)(r+1)^2}$, in

²⁸ If $P_B^*(\frac{k}{\pi_p})$ is not an integer then the optimal is whichever of $\lfloor P_B^*(\frac{k}{\pi_p}) \rfloor$ and $\lceil P_B^*(\frac{k}{\pi_p}) \rceil$ gives higher profit, where in general $\lfloor P_B^*(\cdot) \rfloor = \max\{m \in \mathbb{Z} | m \leq P_B^*(\cdot)\}$ is the floor function of P_B^* and $\lceil P_B^*(\cdot) \rceil = \max\{m \in \mathbb{Z} | m \geq P_B^*(\cdot)\}$ is the ceiling function of P_B^* .

²⁹ Notice that, for $p < 1/2$ the function consists of two branches, in each of which is strictly monotone. Hence, we can define the inverse function in each one of the two branches separately (but not in the whole domain). In this case, given that the interior maximum always satisfies $t^* \geq \frac{n}{2}$, then we define the inverse in the subset $[\frac{n}{2}, n]$ of the domain.

which case $t^* = 0$. For the rest of the cases, it would be helpful to evaluate the derivative at $t = 0$ and $t = n$. Notice that, in this case t and s appear as denominators in parts of the expressions, hence the derivative is not well defined at these points. However we can calculate the limits of the derivative as t tends to 0 and n respectively. Using the formulas from Lemma 5 it is straight forward to find that $\lim_{t \rightarrow 0} P_B^{*'}(t) = \lim_{t \rightarrow n} P_B^{*'}(t) = 0$. Hence, if $k = 0$ the derivative of expected profits is always positive, which yields that $t^* = n$. To the contrary, if $0 < k < \frac{4\pi_p r^2 \ln r}{n(r-1)(r+1)^2}$ the derivative has two roots, t_1 and t_2 , one to the right and one to the left of $n/2$. However, only t_2 is a possible maximum since at t_1 the function attains a minimum. This point is interior and, as mentioned above, if we define the inverse function of the derivative only in the subset $[\frac{n}{2}, n]$ of the domain then this point is equal to $t^* = P_B^{*'}(\frac{k}{\pi_p})$, as long as it is an integer number. As before, if it is not an integer, then we refer the reader to footnote 28.

Concluding one can see that in all the cases we analyzed there exist a lower and upper bound of k , denoted by \underline{k} and \bar{k} , whose values depend on p, n, π_i, π_p , such that the optimal number of targeted agents (i) $t^* = n$, if $k \leq \underline{k}$, (ii) $t^* \in (0, n)$, if $\underline{k} < k < \bar{k}$ and (iii) $t^* = 0$, if $k \geq \bar{k}$.

□

Proof of Remark 1. 1) τ strictly concave in s :

$$\frac{\partial \tau}{\partial s} = \frac{1}{2p-1} \left[\frac{1}{2} - \frac{n}{4} \frac{r^{\frac{s}{2}} \ln r}{(r^{\frac{n}{2}} - 1)} \right] \Rightarrow \frac{\partial^2 \tau}{\partial s^2} = -\frac{n(\ln r)^2 r^{\frac{s}{2}}}{8(2p-1)(r^{\frac{n}{2}} - 1)} < 0$$

$$\frac{\partial \tau}{\partial t} = \frac{1}{2p-1} \left[\frac{n}{4} \frac{r^{\frac{s}{2}} \ln r}{(r^{\frac{n}{2}} - 1)} - \frac{1}{2} \right] \Rightarrow \frac{\partial^2 \tau}{\partial t^2} = -\frac{n(\ln r)^2 r^{\frac{s}{2}}}{8(2p-1)(r^{\frac{n}{2}} - 1)} < 0$$

2) τ has interior maximum in t , therefore it has interior maximum in s :

We have already found that τ is strictly concave. Hence, we only need to ensure that there exists s^* such that $\frac{\partial \tau}{\partial s} = 0$.

At $s = 0$, $\frac{\partial \tau}{\partial s} = \frac{1}{2p-1} \left[\frac{1}{2} - \frac{n}{4} \frac{\ln r}{(r^{\frac{n}{2}} - 1)} \right] > 0$. This holds because $\frac{n \ln r}{2(r^{\frac{n}{2}} - 1)} < 1 \Leftrightarrow n \ln r - 2r^{\frac{n}{2}} + 2 < 0$, which is true for $r \geq 1$ because this latest expression, with respect to r , attains a unique maximum equal to zero for $r = 1$. Hence, it is strictly negative for all $r > 1$.

At $s = n$, $\frac{\partial \tau}{\partial s} = \frac{1}{2p-1} \left[\frac{1}{2} - \frac{n}{4} \frac{r^{\frac{n}{2}} \ln r}{(r^{\frac{n}{2}} - 1)} \right] < 0$. This holds because $\frac{nr^{\frac{n}{2}} \ln r}{2(r^{\frac{n}{2}} - 1)} > 1 \Leftrightarrow nr^{\frac{n}{2}} \ln r - 2r^{\frac{n}{2}} + 2 > 0$, which is true because this latest expression attains unique minimum equal to zero for $r = 1$. Hence, it is strictly positive for all $r > 1$.

Using the two previous results and the fact that the derivative is continuous in $(0, n)$ for all $r > 1$, we can apply Bolzano theorem and conclude that there exists some s^* such that the derivative becomes equal to zero.

3)

$$\begin{aligned}\lim_{r \rightarrow 1^+} s^* &= 2 \lim_{r \rightarrow 1^+} \frac{\ln(r^{\frac{n}{2}} - 1) - \ln(\ln r) - \ln(\frac{n}{2})}{\ln r} = 2 \lim_{r \rightarrow 1^+} \frac{\frac{n}{2} r^{\frac{n}{2}} \ln r - r^{\frac{n}{2}} + 1}{\ln r (r^{\frac{n}{2}} - 1)} = \\ &= 2 \lim_{r \rightarrow 1^+} \frac{\frac{n^2}{4} r^{\frac{n}{2}} \ln r}{r^{\frac{n}{2}} - 1 + \frac{n}{2} r^{\frac{n}{2}} \ln r} = \frac{n^2}{2} \lim_{r \rightarrow 1^+} \frac{\frac{n}{2} \ln r + 1}{n + \frac{n^2}{4} \ln r} = \frac{n}{2} \\ \lim_{r \rightarrow 1^+} \tau_{(t=\frac{n}{2}, s=\frac{n}{2})} &= \lim_{r \rightarrow 1^+} \frac{1}{2p-1} \left[\frac{n}{4} - \frac{n}{2} \frac{1}{r^{\frac{n}{4}} + 1} \right] \stackrel{0}{=} \lim_{p \rightarrow \frac{1}{2}^+} \frac{n^2 r^{\frac{n}{4}-1}}{16(r^{\frac{n}{4}} + 1)^2} \frac{\partial r}{\partial p} = \frac{n^2}{16}\end{aligned}$$

4)

$$\begin{aligned}\lim_{r \rightarrow +\infty} s^* &= 2 \lim_{r \rightarrow +\infty} \frac{\ln(r^{\frac{n}{2}} - 1) - \ln(\ln r) - \ln(\frac{n}{2})}{\ln r} = 2 \lim_{r \rightarrow +\infty} \left[\frac{nr^{\frac{n}{2}}}{2(r^{\frac{n}{2}} - 1)} - \frac{1}{\ln r} \right] = n \\ \lim_{r \rightarrow +\infty} \tau_{(t=2, s=n-2)} &= \lim_{r \rightarrow +\infty} \frac{1}{2p-1} \left[\frac{n}{2} - 1 - \frac{n}{2} \frac{r^{\frac{n}{2}-1} - 1}{r^{\frac{n}{2}} - 1} \right] = \frac{n}{2} - 1\end{aligned}$$

□

Proof of Remark 2. 1) Call $s' = \frac{s}{n-s}$, $n' = \frac{n}{n-s}$ and notice that $\frac{\partial s'}{\partial s} = \frac{\partial n'}{\partial s} = \frac{n}{(n-s)^2}$. Now,

$$\begin{aligned}\frac{\partial \tau}{\partial s} &= \frac{1}{2p-1} \left[\frac{\partial s'}{\partial s} - \frac{\partial n'}{\partial s} \left(\frac{r^{s'} - 1}{r^{n'} - 1} \right) - n' \frac{\frac{\partial s'}{\partial s} r^{s'} \ln r (r^{n'} - 1) - \frac{\partial n'}{\partial s} r^{n'} \ln r (r^{s'} - 1)}{(r^{n'} - 1)^2} \right] = \\ &= - \frac{n(r^{n'} - r^{s'})}{(2p-1)(n-s)^2(r^{n'} - 1)^2} (r^{n'} - 1 - n' \ln r) < 0\end{aligned}$$

Given that the last term is positive for $r < 1$. The argument for t follows directly, since $t = n - s \Rightarrow \frac{\partial \tau}{\partial t} = -\frac{\partial \tau}{\partial s}$.

2) For $r \rightarrow 0^+$ the calculation is straightforward. For $r \rightarrow 1^-$ we use the previous definition of s' and n' . It is easier to calculate the limit with respect to p as it goes to $\frac{1}{2}$, and applying three times L'Hopital's rule we get

$$\lim_{p \rightarrow \frac{1}{2}} \tau = - \frac{n'}{2} \frac{\frac{1}{(1/2)^2} [s' n' (s' - n')]}{2(1/2)^2 n'^2 \frac{1}{(1/2)^2}} = s' = \frac{s}{n-s}$$

3) The first result is straightforward. For the second one we use again the definition of s' and n' . Notice that $s' = n' - 1$ and take the limit with respect to n' . When $s \rightarrow n$, then $n' \rightarrow \infty$. Hence,

$$\lim_{s \rightarrow n} \tau = \lim_{n' \rightarrow \infty} \frac{1}{2p-1} \left(n' - 1 - n' \frac{r^{n'-1} - 1}{r^{n'} - 1} \right) = \lim_{n' \rightarrow \infty} \frac{1}{2p-1} \left[n' \frac{r^{n'} (r-1)}{r^{n'} - 1} - 1 \right] = 0$$

Because,

$$\lim_{n' \rightarrow \infty} \frac{n' r^{n'-1}}{r^{n'} - 1} = \lim_{n' \rightarrow \infty} \frac{n'}{r - r^{1-n'}} = \lim_{n' \rightarrow \infty} \frac{1}{(1-n') r^{-n'}} = 0$$

□

Proof of Remark 3. 1)

$$\frac{\partial \tau}{\partial s} = \frac{n(r-1)(r^{\frac{n}{s}} - 1 - \frac{n}{s} r^{\frac{n}{s}} \ln r)}{(2p-1)s^2(r^{\frac{n}{s}} - 1)^2} < 0$$

Because the last term is negative for all $\frac{n}{s} > 0$. The argument for t follows directly.

2) For $r \rightarrow 0$ the proof is straightforward by substitution. For $r \rightarrow 1$ is easier to calculate the limit with respect to p , as $p \rightarrow \frac{1}{2}$ and by applying three times L'Hopital rule we get:

$$\lim_{r \rightarrow \frac{1}{2}} \frac{1}{2p-1} \left(1 - \frac{n}{s} \frac{(r-1)}{(r^{\frac{n}{s}} - 1)} \right) = -\frac{n}{2s} \frac{\frac{n}{s}(\frac{n}{s} - 1) - \frac{n}{s}(\frac{n}{s} - 1) + \frac{n}{s}(\frac{n}{s} - 1)(\frac{n}{s} - 2)}{2(\frac{1}{2})^2(\frac{n}{s})^2} = \frac{n}{s} - 1 = \frac{t}{s}$$

2) Remembering that $r = \frac{p}{1-p}$, both proofs are straightforward by substitution.

□

Proof of Proposition 8.

$$\frac{\partial \tau_{(\frac{t}{2m}, \frac{s}{2m})}}{\partial m} = \frac{1}{2p-1} \left[-\frac{s}{2m^2} + \frac{n}{2m^2} \left(\frac{r^{\frac{s}{2m}} - 1}{r^{\frac{n}{2m}} - 1} \right) - \frac{n}{2m} \frac{\partial \left(\frac{r^{\frac{s}{2m}} - 1}{r^{\frac{n}{2m}} - 1} \right)}{\partial m} \right] < 0 \text{ for all } r \neq 1$$

Which holds because of the following:

- The first part of the expression inside the brackets is

$$-\frac{s}{2m^2} + \frac{n}{2m^2} \left(\frac{r^{\frac{s}{2m}} - 1}{r^{\frac{n}{2m}} - 1} \right) = \frac{r^{\frac{s}{2m}} - 1}{m} \left(\frac{n}{2m} \frac{1}{r^{\frac{n}{2m}} - 1} - \frac{s}{2m} \frac{1}{r^{\frac{s}{2m}} - 1} \right) \begin{cases} < 0 & \text{if } r > 1 \\ > 0 & \text{if } r < 1 \end{cases}$$

To show this we need to define the function $f(x) = \frac{x}{r^x - 1}$. Its derivative is $\frac{df}{dx} = \frac{r^x - 1 - r^x x \ln r}{(r^x - 1)^2}$, whose denominator is positive for all $r \neq 1$, whereas the nominator is negative, because $\frac{\partial(r^x - 1 - r^x x \ln r)}{\partial r} = -x^2 r^{x-1} \ln r$ which is positive for $r < 1$ and negative for $r > 1$, obtaining a maximum equal to zero for $r = 1$. Hence, f is strictly decreasing in x for all $r \neq 1$, which makes the expression inside the parenthesis always negative (because $\frac{n}{2m} > \frac{s}{2m}$) and the sign depending only on $r^{\frac{s}{2m}} - 1$ which is positive for $r > 1$ and negative for $r < 1$.

- Now, to find the sign of $-\frac{n}{2m} \frac{\partial \left(\frac{r^{\frac{s}{2m}} - 1}{r^{\frac{n}{2m}} - 1} \right)}{\partial m}$, notice that $\frac{r^{\frac{s}{2m}} - 1}{r^{\frac{n}{2m}} - 1} = 1 - \frac{r^{\frac{n}{2m}} - r^{\frac{s}{2m}}}{r^{\frac{n}{2m}} - 1}$, hence:

$$-\frac{n}{2m} \frac{\partial \left(\frac{r^{\frac{s}{2m}} - 1}{r^{\frac{n}{2m}} - 1} \right)}{\partial m} = \frac{n}{2m} \frac{\partial \left(\frac{r^{\frac{n}{2m}} - r^{\frac{s}{2m}}}{r^{\frac{n}{2m}} - 1} \right)}{\partial m} \begin{cases} < 0 & \text{if } r > 1 \\ > 0 & \text{if } r < 1 \end{cases}$$

which holds by Proposition 1.

Therefore the whole expression inside the brackets is negative for $r > 1$ and positive for $r < 1$. This expression is multiplied by $(2p-1)$ which positive for $r > 1$ and negative for $r < 1$. This makes the derivative of τ with respect to m always negative.

□

Proof of Proposition 9. Like before, although we are interested only in integer values of s_1, s_2, t and n , this is a well-behaving smooth function for all $r \neq 1$. Hence, we can differentiate it with respect to s_1 .

$$\begin{aligned}
\frac{dP}{ds_1} &= \frac{r^n \left[r^{\frac{t}{2}} - 1 \right]}{r^n - 1} \left[\frac{(r^{s_1} \ln r + r^{-s_1-t} \ln r) \left(r^{s_1+\frac{t}{2}} - 1 \right) - (r^{s_1} - r^{-s_1-t}) \left(r^{s_1+\frac{t}{2}} \ln r \right)}{\left(r^{s_1+\frac{t}{2}} - 1 \right)^2} \right] = \\
&= \frac{r^n \left(r^{\frac{t}{2}} - 1 \right) \ln r}{(r^n - 1) \left(r^{s_1+\frac{t}{2}} - 1 \right)^2} \left[r^{2s_1+\frac{t}{2}} + r^{-\frac{t}{2}} - r^{s_1} - r^{-s_1-t} - r^{2s_1+\frac{t}{2}} + r^{-\frac{t}{2}} \right] = \\
&= \frac{r^n \left(r^{\frac{t}{2}} - 1 \right) \ln r}{(r^n - 1) \left(r^{s_1+\frac{t}{2}} - 1 \right)^2} \left(2r^{-\frac{t}{2}} - r^{s_1} - r^{-s_1-t} \right) = \\
&= \frac{r^n \left(r^{\frac{t}{2}} - 1 \right) \ln r}{(r^n - 1) \left(r^{s_1+\frac{t}{2}} - 1 \right)^2} \left[-r^{s_1} \left(r^{2s_1} - 2r^{s_1-\frac{t}{2}} + r^{-t} \right) \right] = \\
&= -\frac{r^n \left(r^{\frac{t}{2}} - 1 \right) \ln r}{(r^n - 1) \left(r^{s_1+\frac{t}{2}} - 1 \right)^2} \left(r^{s_1} - r^{-\frac{t}{2}} \right)^2 r^{-s_1} = -\frac{r^{s_2} \left(r^{\frac{t}{2}} - 1 \right) \ln r}{r^n - 1}
\end{aligned}$$

If $r > 1$, then $\frac{dP}{ds_1} < 0$, so the optimal targeting decision is $s_1 = 0$, i.e. target one corner. Whilst, if $r < 1$, then $\frac{dP}{ds_1} > 0$ and recalling that $s_1 \leq s_2$, the optimal decision is $s_1 = s_2$ for s even, or $s_1 = s_2 - 1$ for s odd, i.e. to target the middle of the line. See also the following figures (Figure 2.26).

For the case of $r = 1$, it is apparent that the $P(B|\cdot)$ does not depend on s_1 , hence every decision yields the same result.

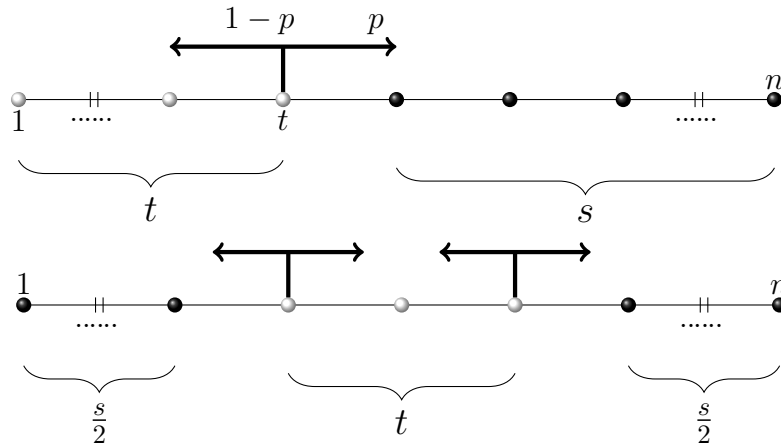


Figure 2.26: Optimal choice for $p > 1/2$ (above) and for $p < 1/2$ (below)

□

Proof of Proposition 10.

$$\begin{aligned}
\frac{dP}{ds_1} &= \frac{r^n \left(r^{\frac{t+1}{2}} - 1 \right)}{r^n - 1} \left[\frac{(r^{s_1} \ln r + r^{-s_1-t} \ln r) \left(r^{s_1+\frac{t+1}{2}} - 1 \right) - (r^{s_1} - r^{-s_1-t}) r^{s_1+\frac{t+1}{2}} \ln r}{\left(r^{s_1+\frac{t+1}{2}} - 1 \right)^2} \right] = \\
&= \frac{r^n \left(r^{\frac{t+1}{2}} - 1 \right) \ln r}{(r^n - 1) \left(r^{s_1+\frac{t+1}{2}} - 1 \right)^2} \left[r^{2s_1+\frac{t+1}{2}} + r^{\frac{-t+1}{2}} - r^{s_1} - r^{-s_1-t} - r^{2s_1+\frac{t+1}{2}} + r^{\frac{-t+1}{2}} \right] = \\
&= \frac{r^n \left(r^{\frac{t+1}{2}} - 1 \right) \ln r}{(r^n - 1) \left(r^{s_1+\frac{t+1}{2}} - 1 \right)^2} \left(2r^{\frac{1-t}{2}} - r^{s_1} - r^{-s_1-t} \right)
\end{aligned}$$

So, the sign of derivative will depend on the sign of $2r^{\frac{1-t}{2}} - r^{s_1} - r^{-s_1-t}$.

For $r < 1$, we can rewrite it as $r^{-s_1} (2r^{\frac{1}{2}} r^{s_1} r^{-\frac{t}{2}} - r^{2s_1} - r^{-t})$ which is negative because:

$$\begin{aligned}
r < 1 &\Rightarrow r^{\frac{1}{2}} < 1 \Rightarrow 2r^{\frac{1}{2}} r^{s_1} r^{-\frac{t}{2}} < r^{s_1} r^{-\frac{t}{2}} \Rightarrow \\
&\Rightarrow 2r^{\frac{1}{2}} r^{s_1} r^{-\frac{t}{2}} - r^{2s_1} - r^{-t} < r^{s_1} r^{-\frac{t}{2}} - r^{2s_1} - r^{-t} \Rightarrow \\
&\Rightarrow 2r^{\frac{1}{2}} r^{s_1} r^{-\frac{t}{2}} - r^{2s_1} - r^{-t} < - \left(r^{s_1} - r^{-\frac{t}{2}} \right)^2 < 0
\end{aligned}$$

So in general for $r < 1$ we get $\frac{dP}{ds_1} > 0$, hence, as before, the optimal decision is $s_1 = s_2$ for s even, or $s_1 = s_2 - 1$ for s odd, i.e. to target the middle of the line.

For $r > 1$, we can rewrite it as $r^{-s_1} r^{-t} \left(2r^{s_1} r^{\frac{1+t}{2}} - r^{2s_1} r^t - 1 \right)$ and naming $x = r^{s_1}$ the content of the parenthesis becomes a polynomial of degree 2, namely $-[r^t x^2 - 2r^{\frac{1+t}{2}} x + 1]$. Let us first calculate the roots of the polynomial which are

$$\begin{aligned}
x_{1,2} &= \frac{2r^{\frac{1+t}{2}} \pm \sqrt{4r^{1+t} - 4r^t}}{2r^t} = r^{\frac{1-t}{2}} \pm (r-1)^{\frac{1}{2}} r^{-\frac{t}{2}} \\
\Rightarrow x_1 &= \left[r^{\frac{1}{2}} - (r-1)^{\frac{1}{2}} \right] r^{-\frac{t}{2}} \text{ and} \\
\Rightarrow x_2 &= \left[r^{\frac{1}{2}} + (r-1)^{\frac{1}{2}} \right] r^{-\frac{t}{2}} \\
- \left(r^t x^2 - 2r^{\frac{1+t}{2}} x + 1 \right) &\geq 0 \Leftrightarrow \left[r^{\frac{1}{2}} - (r-1)^{\frac{1}{2}} \right] r^{-\frac{t}{2}} \leq r^{s_1} \leq \left[r^{\frac{1}{2}} + (r-1)^{\frac{1}{2}} \right] r^{-\frac{t}{2}} \\
&\Leftrightarrow r^{\frac{1}{2}} - (r-1)^{\frac{1}{2}} \leq r^{s_1+\frac{t}{2}} \leq r^{\frac{1}{2}} + (r-1)^{\frac{1}{2}} \\
&\Leftrightarrow \ln r^{s_1} \leq \ln \left(\frac{r^{\frac{1}{2}} - (r-1)^{\frac{1}{2}}}{r^{\frac{t}{2}}} \right) \\
&\Leftrightarrow s_1 \leq \frac{\ln \left[r^{\frac{1}{2}} + (r-1)^{\frac{1}{2}} \right]}{\ln r} - \frac{t}{2} = g(r, t)
\end{aligned}$$

The left hand-side is always satisfied because $r^{\frac{1}{2}} - (r-1)^{\frac{1}{2}} < r^{\frac{1}{2}} < r^{s_1+\frac{t}{2}}$.

So, for $s_1 \leq \frac{\ln[r^{\frac{1}{2}} + (r-1)^{\frac{1}{2}}]}{\ln r} - \frac{t}{2} = g(r, t)$ we find that $\frac{dP}{ds_1} > 0$, while for $s_1 \leq \frac{\ln[r^{\frac{1}{2}} + (r-1)^{\frac{1}{2}}]}{\ln r} - \frac{t}{2} = g(r, t)$ we find $\frac{dP}{ds_1} < 0$, hence the function has a global maximum at $s_1 = \frac{\ln[r^{1/2} + (r-1)^{1/2}]}{\ln r} - \frac{t}{2} = g(r, t)$, however, notice that in our problem s_1 has to be an integer, so in order to find the maximum we need to compare the two closest integers to s_1 , namely $\lfloor g(r, t) \rfloor$ and $\lceil g(r, t) \rceil$. If $P_B(s_1 = \lfloor g(r, t) \rfloor) > P_B(s_1 = \lceil g(r, t) \rceil)$ then $s_1^* = \lfloor g(r, t) \rfloor$ and vice versa.

For $r = 1$, we have that $P_B(s_1; s, t, n, r = 1) = \frac{(t+1)(t+2s_1)}{(t+2s_1+1)n}$ so,

$$\frac{dP}{ds_1} = \frac{t+1}{n} \frac{2(2s_1+t+1) - 2(2s_1+t)}{(2s_1+t+1)^2} = \frac{2(t+1)}{n(2s_1+t+1)^2} > 0$$

Hence, the optimal is to target the middle. \square

Proof of Corollary 2. The proof of Corollary 2 comes directly after the proof of the following lemma (Lemma 8). \square

Lemma 8. *The function $f(r) = \frac{\ln[r^{\frac{1}{2}} + (r-1)^{\frac{1}{2}}]}{\ln r}$ has the following properties.*

- i. $\lim_{r \rightarrow 1^+} f(r) = +\infty$,
- ii. $\lim_{r \rightarrow +\infty} f(r) = \frac{1}{2}$,
- iii. f is strictly decreasing in r ,

Proof of Lemma 8.

$$\begin{aligned} (1) \lim_{r \rightarrow 1^+} f(r) &= \lim_{r \rightarrow 1^+} \frac{\ln[r^{\frac{1}{2}} + (r-1)^{\frac{1}{2}}]}{\ln r} = \lim_{r \rightarrow 1^+} \frac{\frac{1}{r^{\frac{1}{2}} + (r-1)^{\frac{1}{2}}} \left(\frac{1}{2r^{\frac{1}{2}}} + \frac{1}{2(r-1)^{\frac{1}{2}}} \right)}{\frac{1}{r}} = \\ &= \lim_{r \rightarrow 1^+} \left\{ \frac{r}{2r^{\frac{1}{2}} [r^{\frac{1}{2}} + (r-1)^{\frac{1}{2}}]} + \frac{r}{2(r-1)^{\frac{1}{2}} [r^{\frac{1}{2}} + (r-1)^{\frac{1}{2}}]} \right\} = \\ &= \lim_{r \rightarrow 1^+} \frac{r + r \left\{ 2r^{\frac{1}{2}} [r^{\frac{1}{2}} + (r-1)^{\frac{1}{2}}] \right\}}{4r^{\frac{1}{2}} (r-1)^{\frac{1}{2}} [r^{\frac{1}{2}} + (r-1)^{\frac{1}{2}}]^2} = \frac{3}{0^+} = +\infty \\ (2) \lim_{r \rightarrow +\infty} f(r) &= \lim_{r \rightarrow +\infty} \frac{\ln[r^{\frac{1}{2}} + (r-1)^{\frac{1}{2}}]}{\ln r} = \lim_{r \rightarrow +\infty} \frac{\frac{1}{r^{\frac{1}{2}} + (r-1)^{\frac{1}{2}}} \left(\frac{1}{2r^{\frac{1}{2}}} + \frac{1}{2(r-1)^{\frac{1}{2}}} \right)}{\frac{1}{r}} = \\ &= \frac{1}{2} \lim_{r \rightarrow +\infty} \left[\frac{r}{r + (r^2 - r)^{1/2}} + \frac{r}{r + (r^2 - r)^{1/2} - 1} \right] = \\ &= \frac{1}{2} \lim_{r \rightarrow +\infty} \left[\frac{1}{1 + \frac{1}{2(r^2 - r)^{1/2}(2r-1)}} + \frac{1}{1 + \frac{1}{2(r^2 - r)^{1/2}(2r-1)}} \right] = \frac{1}{2} \end{aligned}$$

$$\begin{aligned}
(3) \frac{df}{dr} &= \frac{d\left(\frac{\ln(r^{\frac{1}{2}} + (r-1)^{\frac{1}{2}})}{\ln r}\right)}{dr} = \frac{\frac{\ln r}{r^{\frac{1}{2}} + (r-1)^{\frac{1}{2}}} \left[\frac{1}{2r^{\frac{1}{2}}} + \frac{1}{2(r-1)^{\frac{1}{2}}}\right] - \frac{\ln(r^{\frac{1}{2}} + (r-1)^{\frac{1}{2}})}{r}}{(\ln r)^2} = \\
&= \frac{\left[r^{\frac{1}{2}} + (r-1)^{\frac{1}{2}}\right]}{2\left[r^{\frac{1}{2}} + (r-1)^{\frac{1}{2}}\right] r^{\frac{1}{2}} (r-1)^{\frac{1}{2}} \ln r} - \frac{\ln(r^{\frac{1}{2}} + (r-1)^{\frac{1}{2}})}{r (\ln r)^2} = \\
&= \frac{1}{2r^{\frac{1}{2}} (r-1)^{\frac{1}{2}} \ln r} - \frac{\ln(r^{\frac{1}{2}} + (r-1)^{\frac{1}{2}})}{r (\ln r)^2} < \\
&< \frac{1}{2 \ln r (r-1)} - \frac{\ln(r^{\frac{1}{2}} + (r-1)^{\frac{1}{2}})}{r (\ln r)^2} < \\
&< \frac{1}{2 \ln r (r-1)} - \frac{\ln(2r^{\frac{1}{2}})}{r (\ln r)^2} = \frac{1}{2 \ln r (r-1)} - \frac{\ln 2}{r (\ln r)^2} - \frac{1}{2r \ln r} = \\
&= \frac{1}{2r (r-1) (\ln r)^2} [\ln r - 2 \ln 2 (r-1)] < 0
\end{aligned}$$

because both $[\ln r]$ and $[2 \ln 2 (r-1)]$ are increasing functions, equal to zero for $r = 1$ but the second one increases at a higher rate, because $\frac{d(\ln r)}{dr} = \frac{1}{r} < 2 \ln 2 = \frac{d[2 \ln 2 (r-1)]}{dr}$. Hence, f is decreasing in r . \square

2.9 Appendix C - Figures

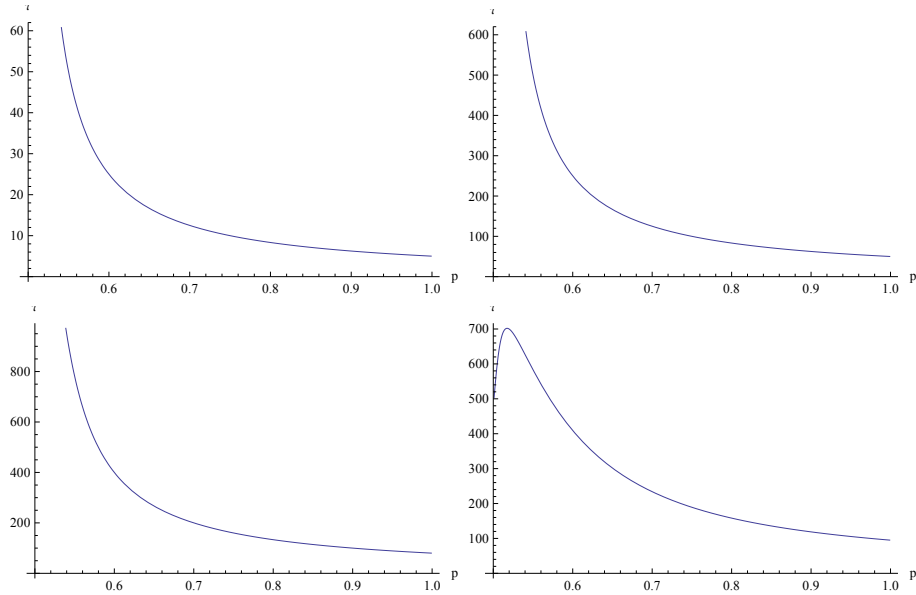


Figure 2.27: $\tau(p)$ for $r > 1$, $n = 200$ and (i) $t = 190$, (ii) $t = 100$, (iii) $t = 40$, (iv) $t = 10$

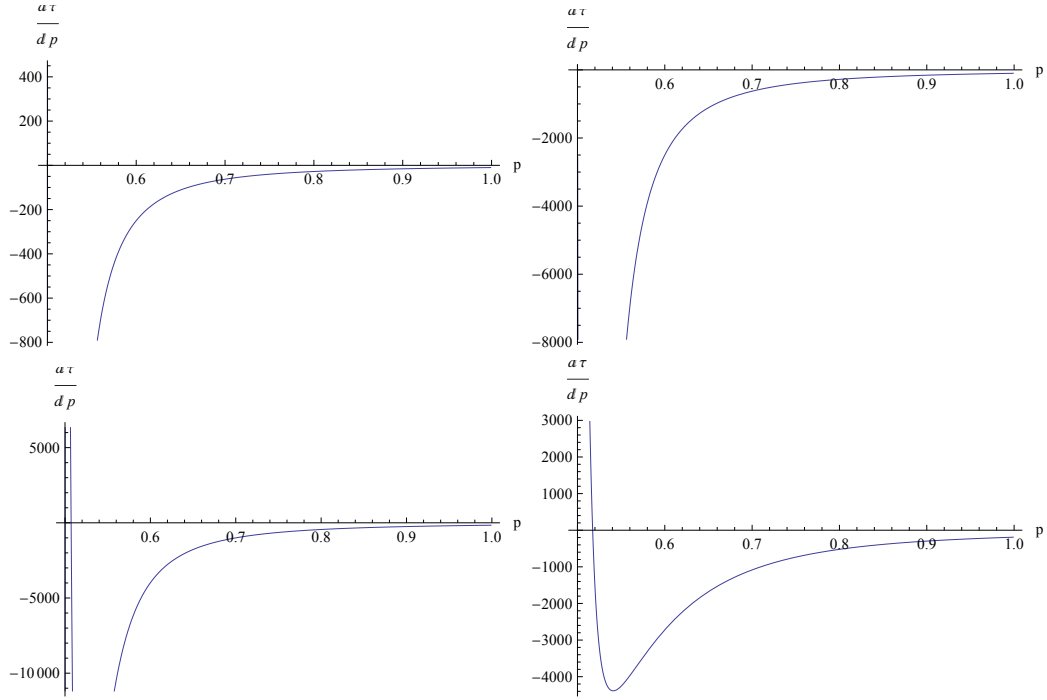


Figure 2.28: $\frac{d\tau(p)}{dp}$ for $r > 1$, $n = 200$ and (i) $t = 190$, (ii) $t = 100$, (iii) $t = 40$, (iv) $t = 10$

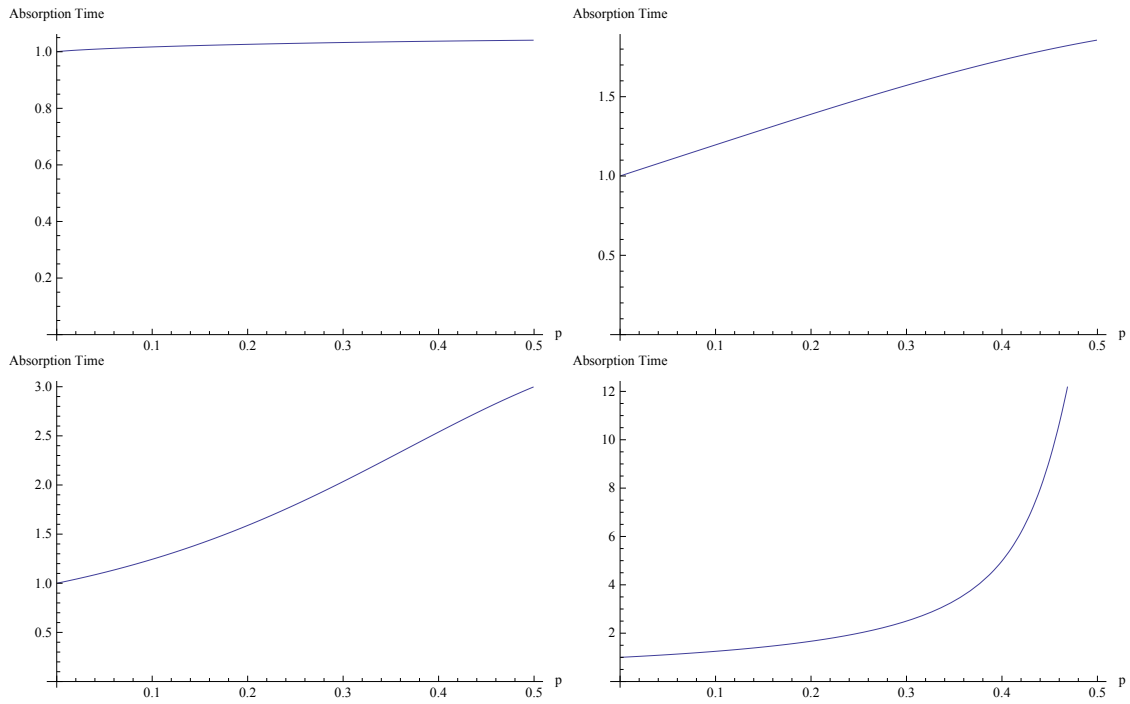


Figure 2.29: $\tau(p)$ for $r < 1$, $t < s$, $n = 200$ and (i) $t = 98$, (ii) $t = 70$, (iii) $t = 50$, (iv) $t = 10$

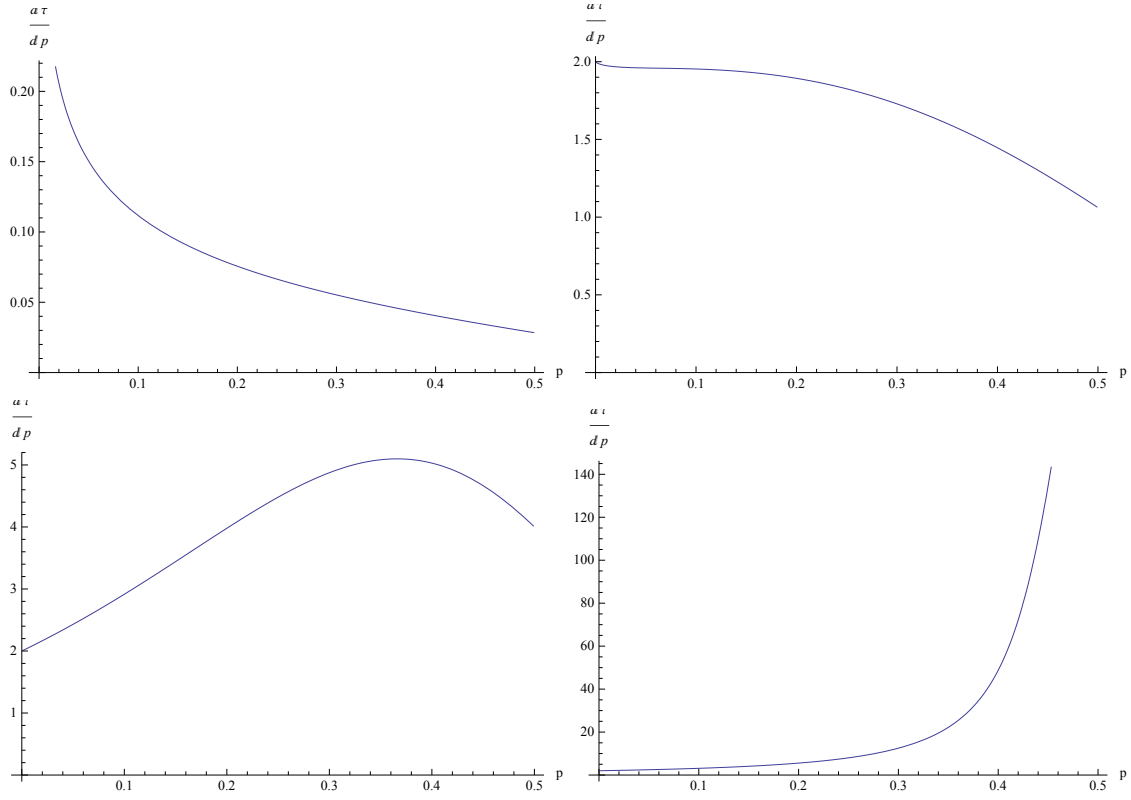


Figure 2.30: $\frac{d\tau(p)}{dp}$ for $r < 1$, $t < s$, $n = 200$ and (i) $t = 98$, (ii) $t = 70$, (iii) $t = 50$, (iv) $t = 10$

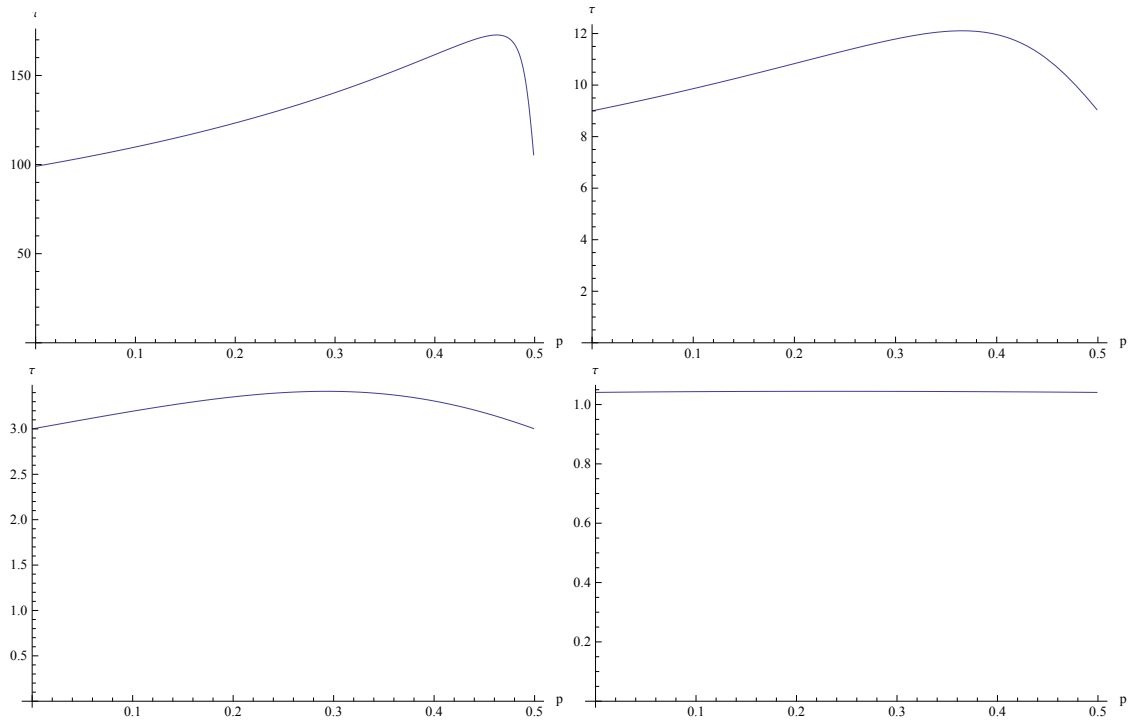


Figure 2.31: $\tau(p)$ for $r < 1$, $t > s$, $n = 200$ and (i) $t = 198$, (ii) $t = 180$, (iii) $t = 150$, (iv) $t = 102$

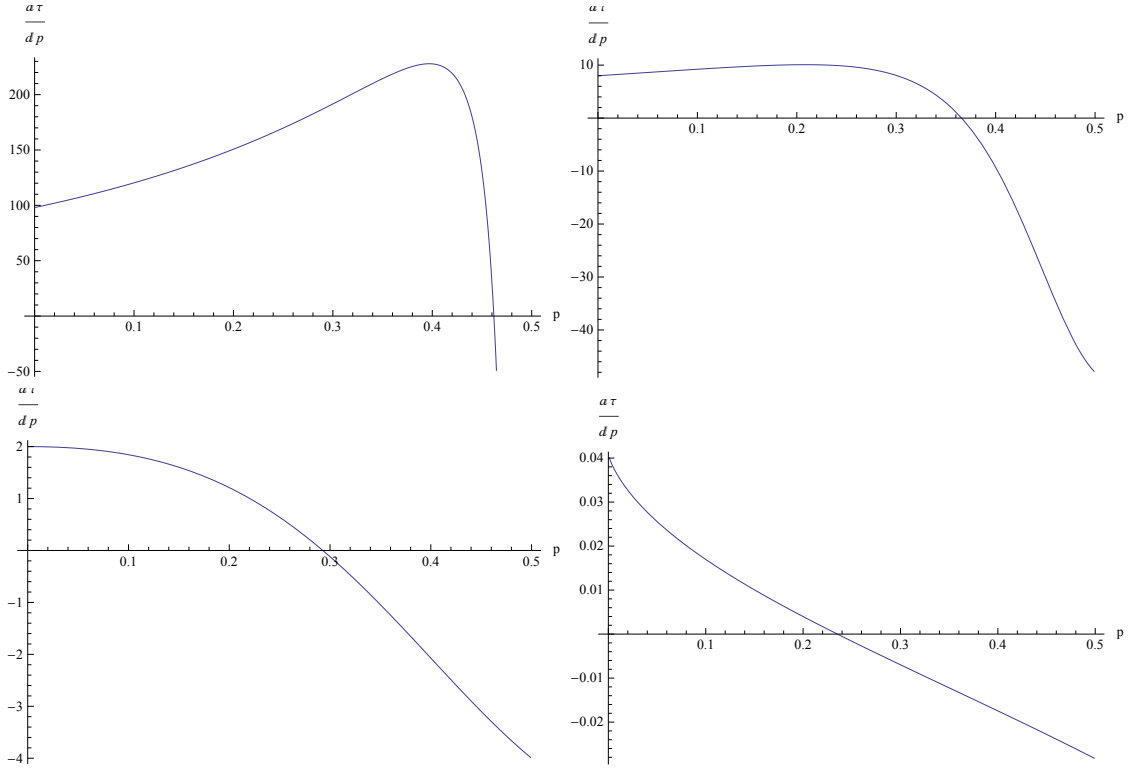


Figure 2.32: $\frac{d\tau(p)}{dp}$ for $r < 1$, $t > s$, $n = 200$ and (i) $t = 198$, (ii) $t = 180$, (iii) $t = 150$, (iv) $t = 102$

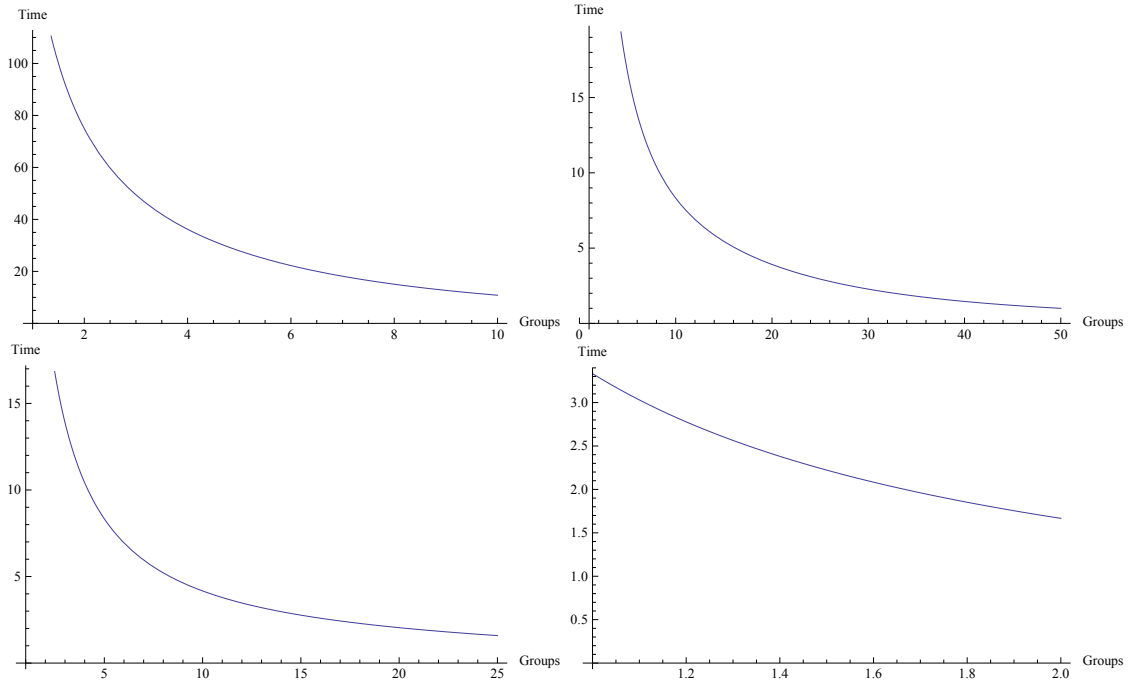


Figure 2.33: $\tau(m)$ for $r = 0.2$, $n = 200$ and (i) $t = 180$, (ii) $t = 100$, (iii) $t = 50$, (iv) $t = 4$

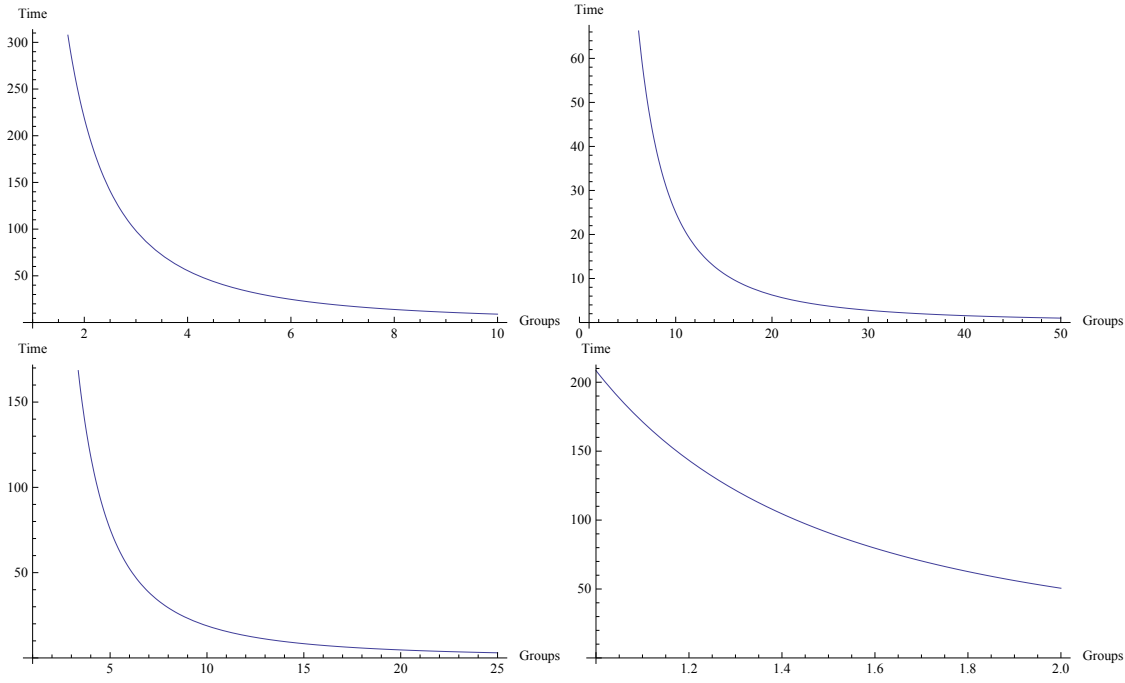


Figure 2.34: $\tau(m)$ for $r = 0.501$, $n = 200$ and (i) $t = 180$, (ii) $t = 100$, (iii) $t = 50$, (iv) $t = 4$

2.10 Appendix D - Matlab Code

```

1 startLoop = tic; % Time counter
2 n=20; % Population
3 t=4; % Targeted agents
4 repetitions=4000; % Number of repetitions to approximate p
5 % Grid of values of p, which will be 0.15,0.25,...,0.85
6 pgrid=8;
7 % Add a link between agent 1 and agent "connect"
8 connect=3;
9
10 A = eye(n); % Define the adjacency matrix
11 for ai = 2:(n-1)
12     A(ai, ai - 1) = 1;
13     A(ai, ai + 1) = 1;
14 end
15 A(1, n) = 1;
16 A(1, 2) = 1;
17 A(n, 1) = 1;
18 A(n, n-1) = 1;
19
20 A(1, connect) = 1;
21 A(connect, 1) = 1;
22
23 % Define possible initial configurations for t=4 (under A1)
24 S = [];

```

```

25
26 for v1 = 1:2:17
27     for v2 = v1+2:2:19
28         v3 = [v1,v1+1,v2,v2+1];
29         S = [S; v3];
30     end
31 end
32
33 for v4 = 2:2:16
34     for v5 = v4+2:2:18
35         v6=[v4,v4+1,v5,v5+1];
36         S=[S; v6];
37     end
38     v7=[v4,v4+1,20,1];
39     S=[S; v7];
40 end
41
42 v8=[18,19,20,1];
43 S=[S;v8];
44
45 [b,my]=size(S)
46
47
48 % Configurations as arrays of binary variables
49 Config = zeros(b,n);
50 for i = 1:b
51     for j = 1:t
52         Config(i, S(i, j)) = 1;
53     end
54 end
55 % Probability of successful diffusion
56 % for all configurations and values of p
57 PB = zeros(b,pgrid);
58 % Average time of absorption
59 % for all configurations and values of p
60 AvTime = zeros(b,pgrid);
61
62 % The actual process for each configuration
63 for i = 1:b
64     for j = 1:pgrid
65         s1 = 0;
66         f1 = 0;
67         Periods1 = 0;
68
69         for k1 = 1:repetitions
70 % Matrix of realized configurations in each repetition
71     Choice1 = ones(301,n)*2;

```

```

72     Choice1(1,:) = Config(i,:);
73     t1 = 1;
74 % Checking if convergence has occurred
75     while (isequal(Choice1(t1,:),ones(1,n))==0 && isequal(Choice1(t1,:),zeros(1,n))
           ==0)
76 % Generating the random shock of the period
77 % c_1 is a random number from a Uniform [0,1]
78     c1 = random('Uniform',0,1)
79     for i1 = 1:n
80         for j1 = 1:n
81 %Revision can occur only if two agents are connected
82         if A(i1, j1) == 1
83 % And only if they have chosen differently
84 % in the previous period
85         if Choice1(t1, i1) ~= Choice1(t1, j1)
86 % Given p, there is success if c1>1-p,
87 % for p=0.15,0.25,...,0.85
88         if c1 >= 0.55+0.05*pgrid-j/10
89             Choice1(t1 + 1, i1) = 1;
90         else
91             Choice1(t1 + 1, i1) = 0;
92         end
93     end
94 end
95 end
96 % If no condition for revision has been satisfied ,
97 % choose the same as in the previous period
98     if Choice1(t1 + 1, i1) == 2
99         Choice1(t1 + 1, i1) = Choice1(t1, i1);
100     end
101 end
102     t1 = t1 + 1;
103 end
104 % In case of successful convergence
105     if Choice1(t1, 1) == 1
106         s1 = s1 + 1;
107     else
108 % In case of unsuccessful convergence
109         f1 = f1 + 1;
110     end
111     Periods1 = Periods1 + t1 - 1; % counter of periods
112 end
113 % Update of the estimators of probability and average time
114     PB(i, j) = s1/(s1 + f1);
115     AvTime(i, j) = Periods1/repetitions;
116 end
117 end

```

```

118
119 % Find the optimal configuration and the maximum
120 % probability of successful diffusion for each possible p
121 MaxProb = zeros(pgrid,1);
122 % Assume the first configuration is the optimal
123 MaxProb = PB(1,:);
124 OptimalConfig = ones(1,pgrid);
125 for j = 1:pgrid
126     for i = 2:b
127 % If the current element is larger than the current maximum
128 % substitute it
129         if PB(i, j) > MaxProb(j)
130             MaxProb(j) = PB(i, j);
131 % Substitute also the configuration tha yields the maximum
132             OptimalConfig(j) = i;
133         end
134     end
135 end
136
137 OptimalConfig
138 % Optimal configurations for each p
139 Opt = S(OptimalConfig,:);
140 % Vector of maximum probabilities for each p
141 MaxProb
142
143 endLoop=toc(startLoop); % End time counter

```

Chapter 3 is dedicated to my friends:
Alexi C, Alexi K, Gianni and Kosta.

Chapter 3

Optimal Influence under Observational Learning

3.1 Introduction

3.1.1 Motivation

The role of influential agents in the diffusion of products and ideas has been subject to extended research in several different fields (see for example Domingos and Richardson, 2001; Kempe et al., 2003, 2005; Galeotti and Goyal, 2009; Kirby and Marsden, 2006, and references therein). Recent technological advances have made possible the collection of previously unobservable data related to particular individual characteristics. Based on such observations, researchers, firms and other interested parties are able to identify the structure and the kind of social interactions between members of a society.

Observing these interactions is particularly useful for firms that want to determine effective targeting strategies at the individual level, attempting to maximize the spread of their products in a market. Given that targeting agents is costly, the effective selection of targets becomes crucial. The targeting problem becomes even tougher because the firms are not always aware of the relative (compared to those of their competitors) quality of their product *ex-ante*. This is a commonly observed issue, since it is normal that a firm can determine the level of satisfaction of the consumers only after the product has been circulated in the market.

There is a growing literature dealing with optimal influence in networks, which focuses on different network structures and behavioral rules (see Galeotti and Goyal, 2009; Goyal and Kearns, 2012; Ortuño, 1993). Chatterjee and Dutta (2011) analyze a similar question to ours, however in their framework the firm knows exactly the quality of the product and the agents are either completely naive, in the sense that they adopt the product as soon as they observe it, or fully rational bayesian maximizers. Nevertheless, in their paper one of the most crucial parameters that identify the optimal targeting strategies is the decay centrality of the agents, which turns out to be the case also in this paper. The most closely related paper is by Tsakas (2013), where the author analyzes a similar

problem of targeting agents. The main difference is that uncertainty regarding the quality of the product persists forever, leading the agents to revise their choices repeatedly. This effect leads to significant changes in the optimal targeting strategies.

In this paper, we want to identify the optimal targeting strategy of a firm that is ex-ante uncertain about the relative quality of its product, while the agents can perfectly learn the quality of a product as soon as they have observed someone choosing it. In an environment where agents choose repeatedly between two alternative products, if the product is of higher quality than its alternative, then it will get diffused to the whole population; hence the firm would like diffusion to occur as fast as possible. To the contrary, if the product is of lower quality, then it will survive only for a finite number of periods, and in particular until every agent has observed the alternative at least once. Therefore, the firm would like to protect the early consumers of the product from observing the alternative.

3.1.2 Setting

We consider an interested party, from now on called *the firm*, who seeks to maximize the diffusion of an action, from now on called *product B* in a social network with finitely many agents. The agents choose repeatedly between product *B* and an alternative product *A* and they observe the qualities of the products chosen by themselves and by their immediate neighbors. The firm can target a subset of the population to choose the product in the first period. The objective of the firm is to target the subset that will maximize the lifetime discounted sum of products *B* consumed.

The agents choose the product that they have observed to be of higher quality. The quality of each product remains unchanged throughout the process, therefore the agents may revise their choice only the period after they observed both products for the first time. They do not revise earlier because they are not aware of the alternative and they do not revise later because they already know which product is of higher quality.

The firm does not know ex-ante whether its product is the one of higher quality. However, it has some limited information that determines the probability that this is the case. Therefore, the diffusion of the product is uncertain and the targeting strategy has to be designed taking this factor of uncertainty into account. In practice, this information is related to the relative performance of the product and the level of satisfaction of the consumers, which cannot be known to the firm unless the product has already been circulated.

3.1.3 Results

We provide a necessary and sufficient condition that characterizes the optimal targets for any network structure. More specifically, in the particular case that the firm can target only one agent, then it is optimal to target the agent with the maximum decay centrality. In the general case where the firm can target a larger subset of the population the optimal strategy combines two features: On the one hand, should intend to maximize the decay centrality of the targeted set, so as to capture the population quickly in case the product is of high quality. On the other hand, should seek to minimize the decay

centrality of the targeted set's complement, so that the product survives as long as possible even if it is of low quality. If the firm is extremely impatient, then decay centrality coincides with degree centrality, whereas if the firm is extremely patient then decay centrality coincides with closeness.

Given that decay centrality is a measure that depends crucially on the exact network structure, we study separately the case of the circle as an attempt to provide a more intuitive picture of this measure. In this case, we identify the shape of the targeted subset, which should consist of one large group of connected agents and several others spread uniformly around the network. The exact number of groups depends on the particular characteristics of each environment and for this we provide some partial results and numerical examples. In particular, if the firm is sufficiently optimistic about the product, then prefers to spread it as uniformly as possible around the society, whereas if the firm is sufficiently pessimistic prefers to concentrate all the targeted agents together. This result is true for any level of patience and thus is in partial contrast with the results obtained by Tsakas (2013).

3.2 Model

3.2.1 The Network

There is a finite set of agents N , who are connected through a social network. The set has cardinality n , typical elements i and j and is mentioned as *population* of the network.

A social network is represented by a family of sets $\mathcal{N} := \{N_i \subseteq N \mid i = 1, \dots, n\}$, with N_i denoting the set of agents that can observe and be observed by i . Throughout the paper N_i is called i 's neighborhood and is assumed to contain i ¹. Its cardinality, $|N_i|$, is called i 's degree and the agents who belong to N_i are called i 's neighbors. In the present setting, the network structure describes the channels of communication in the population and does not impose strategic interactions. More specifically, each agent $i \in N$ observes the choices and their associated outcomes of all $j \in N_i$. We focus on undirected networks, where $j \in N_i$ if and only if $i \in N_j$, nevertheless with appropriate modifications the results for directed networks would look very similar.

A path in a network between agents i and j is a sequence i_1, \dots, i_K such that $i_1 = i$, $i_K = j$ and $i_{k+1} \in N_{i_k}$ for $k = 1, \dots, K - 1$. The geodesic distance, $d(i, j)$, between two agents in the network is the length of the shortest path between them. More generally, the geodesic distance, $d(T, j)$, between a set of agents T and an agent outside of the set, $j \in T^C := N \setminus T$,² is defined as the minimum among all the distances between some agent $i \in T$ and agent $j \in T^C$, i.e. $d(T, j) := \min_{i \in T} d(i, j)$.

We say that two agents are connected if there exists a path between them. The network is connected if every pair of agents is connected. We focus on connected networks, nevertheless for disconnected networks the analysis would be identical for each of their connected components.³

¹This assumption is not usual, however in this setting it is necessary since N_i describes the set of agents whose choices can be observed by i . Therefore, since it is reasonable to assume that one can observe her own choices, she should be contained in her own neighborhood.

²Throughout the paper, T^C denotes complement of T .

³A connected component is a non-empty sub-network N' such that (i) $N' \subset N$, (ii) N' is connected and (iii) if

3.2.2 The Problem

Each period ($\tau = 1, \dots$), each agent consumes one unit of either product A , whose quality (q_A) is commonly known, or product B , whose quality is ex-ante uncertain. An interested party, from now on called the *firm*, wants to propagate the consumption of product B to the network. It can do so by convincing a subset of the population, $T \subset N$ with cardinality $t \leq n$, to consume the product once.⁴ The goal of the firm is to target the subset of the population that will maximize the expected total consumption, from now on called *sales*, of the product. We refer to this as the *optimal targeting strategy* of the firm. The firm discounts future sales at a rate $\delta \in (0, 1)$. The product's quality (q_B) is either *high* (H), with probability $p \in [0, 1]$, or *low* (L), with probability $1 - p$. To avoid trivial scenarios, we assume that H and L satisfy $L < q_A < H$. The true quality, $q_B \in \{H, L\}$, is ex-ante unknown to the firm, therefore the optimal targeting strategy will depend on p .

An agent becomes aware of the product either if she is targeted by the firm, or if she observes one of her neighbors consuming it. As soon as she becomes aware, she learns which product is of higher quality and consumes it thereafter.

As it becomes apparent, there are two opposite forces that will be driving the results. On the one hand, if product B 's quality is high then it will certainly be diffused to the whole population after a finite number of periods. Hence, given the presence of discounting, the firm would like to capture the whole population as quickly as possible. On the other hand, if its quality is low, then it will unavoidably disappear after finitely many periods. In this case, the firm would want to make the product survive for as many periods as possible.

To formally define the problem of the firm we need to introduce some more notation. Let \mathcal{T} be the collection of subsets of N with cardinality t and $S_\tau(T|q_B)$ be the firm's sales function, in period τ , for a targeting strategy $T \in \mathcal{T}$, and given that B 's quality is q_B . Therefore, the firm seeks the targeting strategy $T \in \mathcal{T}$ that maximizes the following payoff function:

$$\Pi(T) := E_{q_B} \left[\sum_{\tau=1}^{+\infty} \delta^{\tau-1} S_\tau(T) \right] = p \sum_{\tau=1}^{+\infty} \delta^{\tau-1} S_\tau(T|q_B = H) + (1-p) \sum_{\tau=1}^{+\infty} \delta^{\tau-1} S_\tau(T|q_B = L)$$

3.3 Results: Targeting only one agent

First, we study the case where the firm can target only one agent. In this case, it is apparent that if the quality of the product is low then the payoff of the firm is zero, irrespectively of the targeting choice. Therefore, the firm focuses on maximizing the discounted sum of sales in case the quality of the product is high. The parameter that determines the optimal targeting strategy is the decay centrality of the agents, which is defined as follows (Jackson, 2008):

Definition 5. Consider a decay parameter $\delta \in (0, 1)$. The decay centrality of an agent $i \in N$ is defined as $\sum_{j \neq i} \delta^{d(i,j)}$.

$i, j \in N'$ and $j \in N_i$, then $j \in N'_i$.

⁴For simplicity, we assume that the firm does not intervene again.

In the current setting the decay parameter δ coincides with the discount factor of the firm and $d(i, j)$ is the geodesic distance between agents i and j . It is the same kind of centrality that determines the optimal targeting strategy Chatterjee and Dutta (2011), despite the fact that the settings differ significantly. Intuitively, this is a measure of centrality that takes into account how close an agent is to other agents, but in a way that very distant agents are weighted less than closer ones. In a sense, it captures how quickly a product of high quality can become visible to the whole population. Agents with high decay centrality are expected to be those who combine a higher number of neighbors with a small geodesic distance from the most isolated agents in the network. The following theorem characterizes the optimal strategy for a firm that targets only one agent ($t = 1$).

Theorem 3. *For all $(\delta, p) \in (0, 1)^2$, the optimal strategy for the firm is to target the agent $i \in N$ who maximizes decay centrality.*

All proofs can be found in the Appendix.

Notice that the exact optimal strategy depends on the structure of the network and it is not easily identified for general structures. As it is also mentioned by Chatterjee and Dutta (2011), this is easily identified in simple structures such as the star, where the center is the optimal target, and the line where the median is the one maximizing decay centrality.

An obviously crucial parameter is the value of the discount factor δ . Intuitively, the lower the value of δ the more important becomes the degree of the targeted agent. If δ gets very close to 0, then the firm cares only about the payoff of the first period, which comes from the sales to agents who have geodesic distance equal to one from the targeted agent. Hence, it is optimal to target the agent with the maximum degree. This result is formally stated in the following corollary, whose proof is omitted, since it follows directly from the proof of the theorem. If $\delta = 0$ the firm cares only about the payoff in the first period. Nevertheless, notice that the same result is also true for positive δ as long as it is sufficiently small.

Corollary 3. *If $\delta = 0$, then the optimal strategy for the firm is to target the agent $i \in N$ with the maximum degree, $|N_i|$.*

In the other extreme case where $\delta = 1$ the result is not that obvious. First, observe that the payoffs of the firm become infinitely large irrespectively of the targeting strategy. Nevertheless, considering the difference in payoffs between alternative targeting strategies we can determine which will be the optimal one, despite the fact that these differences will only be finitely large. In order to obtain a formal result we should slightly modify the original objective function of the firm in the following way: Consider that the firm cares equally about the expected sales of the next P periods, i.e. $\Pi(T) = E_{q_B} \left[\sum_{\tau=1}^P S_\tau(T) \right]$, P can be any finite number, as long as it is sufficiently large to ensure total diffusion for some targeting strategy.⁵ As long as P is sufficiently large the difference in payoffs generated by different targeting strategies remains unchanged and for P being finite this difference

⁵For this to happen it suffices that P is larger than the value of the largest geodesic distance between any two nodes in the network, which is usually called diameter of the network.

is not negligible. It turns out that in this case the optimal strategy for the firm is to target the agent with the maximum closeness (or equivalently the minimum farness), which are both formally defined below.

Definition 6. The farness of an agent $i \in N$ is defined as the sum of the geodesic distances from each other agent in the network, i.e. $\sum_{j \neq i} d(i, j)$, and its closeness is defined as the inverse of the farness.

Proposition 11. *If $\delta = 1$, then in the modified problem the optimal strategy for the firm is to target the agent with the maximum closeness (or equivalently with the minimum farness).*

In the following section, the introduced concepts are extended to cases where the firm can target several agents, rather than a single one. As we will see below, the results are slightly modified, but the intuition remains similar. An important feature is that the members of the set that is optimal to be targeted might be different from those who would be optimal as individual targets.

3.4 Results: Targeting several agents

The problem becomes more complicated if the firm targets several agents. In this case, we can no longer ignore the payoffs of the firm when the product is of low quality. This happens because when the firm can target many agents, it can distribute them in such a way that some of them do not get aware of the alternative product directly, because none of their neighbors uses it initially. Before proceeding to the formal statement it is necessary to modify the previous definitions slightly, so as to be adequate for group with multiple agents. The generalized definitions are based on Everett and Borgatti (1999). Recall that, the geodesic distance between a group of agents T and an agent $j \notin T$ is defined as the minimum distance between some agent $i \in T$ and agent j . Namely, $d(T, j) := \min_{i \in T} d(i, j)$. Also, \mathcal{T} denotes the collection of subsets of N with cardinality t and $T^C = N \setminus T$ is the complement of T .

Definition 7. The *group degree* of a set $T \in \mathcal{T}$, denoted by $|N_T|$, is defined as the number of agents not belonging to T who have at least one neighbor in T , or equivalently the cardinality of the union of neighborhoods of all agents in T , i.e. $|N_T| := \left| \bigcup_{i \in T} N_i \right|$.

The importance of an agent's degree lies on the fact that its actions are observed by others. When targeting multiple agents, what is important is the total number of agents that can observe at least one of the targeted agents. Therefore, the parameter of interest is not the sum of neighbors of all targeted agents per se, but the sum of different neighbors who are directly connected with some agent in the targeted set.

Definition 8. The *group farness* of a set $T \in \mathcal{T}$ is defined as the sum of the distances between T and each other agent outside T , i.e. $\sum_{j \in T^C} d(T, j)$. Its *group closeness* is defined as the inverse of the group farness.

The definition is totally analogous to the case of one agent, taking into account that the focus now is on the necessary steps to reach an agent outside the set starting from any agent in the set.

Definition 9. Consider a decay parameter $\delta \in (0, 1)$. The *group decay centrality* of set $T \in \mathcal{T}$ is defined as $\sum_{j \in T^C} \delta^{d(T,j)}$.

This definition is a generalization of decay centrality. Given that the focus is on sets of multiple agents, the important parameters are the lengths of shortest paths that start from some agent inside the set and end at any agent outside it. Intuitively, this means that the aim of a targeting strategy is to affect the number of periods needed before every agent becomes aware of both product. Ideally, this number should be minimized in case the product is of high quality and be maximized otherwise.

Theorem 4. For $(\delta, p) \in (0, 1)^2$, among all $T \in \mathcal{T}$ (subsets of N with t agents) the optimal strategy is to target the set that maximizes the following expression:

$$p \sum_{j \in T^C} \delta^{d(T,j)} - (1 - p) \sum_{i \in T} \delta^{d(T^C,i)}$$

The above expression is very intuitive and the idea is similar to the case with only one targeted agent. In words, the firm seeks to target the set that maximizes a linear combination of two opposite features: On the one hand, the firm wants to maximize the group decay centrality of the targeted set, so as to capture the whole population as quickly as possible in case it is of high quality. On the other hand, it wants to minimize the group decay centrality of the targeted set's complement, so that the product will survive for as many periods as possible in case it is of low quality. The level of importance of each one of the two factors is determined by the probability p of the product being of high quality. If the product is very likely to be good, then the firm cares almost exclusively about minimizing the number of periods needed for the product to be observed by everyone in the network. Whereas, if the product is very likely to be bad, then the firm cares about protecting the targeted agents from observing the alternative products for as many periods as possible. The following corollaries summarize the results for the extreme values of p and δ . The proofs are omitted since they follow directly from the objective function of the firm and the proof of Theorem 4. The trade-off will become even clearer in the following subsection that we analyze the particular case of the circle network.

Corollary 4. For $p = 1$, among all $T \in \mathcal{T}$ (subsets of N with t agents) the optimal strategy for the firm is to target the set:

- with the maximum group degree, if $\delta = 0$.
- with the maximum closeness, if $\delta = 1$.
- with the maximum decay centrality, if $\delta \in (0, 1)$.

Corollary 5. For $p = 0$, among all $T \in \mathcal{T}$ (subsets of N with t agents) the optimal strategy for the firm is to target the set:

- whose complement, T^C , has the minimum group degree, if $\delta = 0$.
- whose complement, T^C , has the minimum closeness, if $\delta = 1$.
- whose complement, T^C , has the minimum decay centrality, if $\delta \in (0, 1)$.

Corollary 6. For $p \in (0, 1)$, among all $T \in \mathcal{T}$ (subsets of N with t agents) the optimal strategy for the firm is to target the set that:

- maximizes $p |N_T| - (1 - p) |N_{T^C}|$, if $\delta = 0$.
- minimizes $p \sum_{j \in T^C} d(T, j) - (1 - p) \sum_{i \in T} d(T^C, i)$, if $\delta = 1$.

3.4.1 The Circle

It has already been mentioned that decay centrality depends much on the exact topology of the network. In this section, we intend to provide an intuitive image of the set that maximizes decay centrality in a simple network structure; the circle.⁶ The fact that all the agents not only have the same degree, but they are also identical in any measure of centrality provides additional interest to the results regarding the targeting of several agents.⁷

Before proceeding to the results we need to introduce some more notation. If the firm targets t agents, then initially the population consists of s agents choosing product A and t agents choosing product B , where $s + t = n$. We call *group* a sequence of neighboring agents who all choose the same product and are surrounded by agents choosing the alternative product. The population is formed of g groups of neighboring agents who choose product A , with sizes $\{s_1, s_2, \dots, s_g\}$, where $\sum_{k=1}^g s_k = s$ and analogously g groups of neighboring agents who all choose product B , with sizes $\{t_1, t_2, \dots, t_g\}$, where $\sum_{k=1}^g t_k = t$.⁸ We mention these groups as being *of type A* and *of type B* respectively. The numbering of the groups is based on their size in increasing order, $s_1 \leq s_2 \leq \dots \leq s_g$ and $t_1 \leq t_2 \leq \dots \leq t_g$. With some abuse of notation we also use s_1, s_2, \dots, s_g and t_1, t_2, \dots, t_g to name the groups. Our goal is to find the optimal size of all s_k and t_k for $k = 1, \dots, g$, their optimal position (if it matters), as well as the optimal number of groups, g . Figure 3.1 shows an example of a possible initial configuration. White nodes represent agents choosing action B and black nodes agents choosing action A .

In order to avoid unnecessary complications in the calculations (which arise without the gain of any additional intuition) we assume that every group must have an *even* number of agents. Formally:

Assumption 3 (A1). s_i and t_i are even numbers for all $i \in \{1, \dots, g\}$.

⁶The circle is a network where each agent has exactly two immediate neighbors; Namely $N_i = \{i - 1, i, i + 1\}$ for $i = 2, \dots, n - 1$, whereas $N_1 = \{n, 1, 2\}$ and $N_n = \{n - 1, n, 1\}$. The current structure imposes an undirected network, because for all $i, j \in N$, $j \in N_i$ if and only if $i \in N_j$.

⁷It is obvious that the decision is trivial when the firm can target only one agent.

⁸Notice that, the fact that the network is a circle and there exist exactly two products ensures that the number of groups is the same for both products.

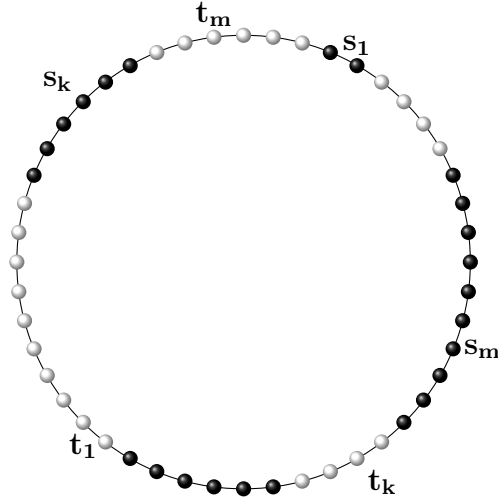


Figure 3.1: This is an example of an initial configuration.

First of all, one should notice that, for a given number of groups, the sizes of groups t_1, \dots, t_g are important only when the product is of low quality, whereas those of s_1, \dots, s_g only when the product is of high quality. This is a direct consequence of Theorem 4 and the fact that all the agents are positionally identical.

Proposition 12. *For a given number of targeted groups, denoted by g , the optimal configuration satisfies the following conditions:*

- i. $s_g - s_1 \leq 2$
- ii. $t_1 = \dots = t_{g-1} = 2$ and $t_g = t - 2(g - 1)$

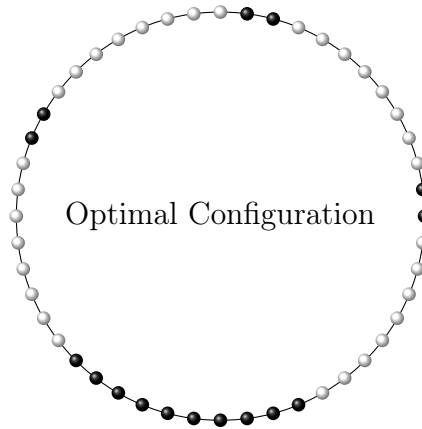


Figure 3.2: The optimal targeting strategy if the firm targets four groups.

The first condition captures the fact that the firm wants to maximize the decay centrality of the targeted set if the product is of high quality. This statement is equivalent to saying that the firm wants to make the product visible to everyone in the population as quickly as possible. Therefore, it

locates the targeted agents in a way such that all groups of type B have almost the same size.⁹ This argument is captured by the first condition.

The second condition captures the fact that the firm wants to minimize the decay centrality of the complement of the targeted set if the product is of low quality. That is equivalent saying that it wants to maximize the number of periods until all the targeted agents have observed the alternative product. The following figure depicts a typical configuration when the number of targeted groups is fixed to four.

However, the above result assumes the number of groups to be fixed, which need not be the case. It is still necessary to compare configurations that satisfy the above conditions and consist of different number of groups. The following two propositions, accompanied by Figures 3.3 and 3.4 intend to clarify this comparison.

Proposition 13. *Assume that the firm can only target either one or two groups and that s is a multiple of 4.¹⁰ Then for any $(p, \delta) \in (0, 1)^2$ it is optimal to target two groups if and only if*

$$p \geq \frac{(1-\delta)(1-\delta^{\frac{t}{2}-1})}{(1-\delta)(1-\delta^{\frac{t}{2}-1}) + (1-\delta^{\frac{s}{4}})^2}.$$

Proposition 13 states that for any level of patience a sufficiently optimistic firm (with p high enough) would prefer to spread the targeted agents, whereas a sufficiently pessimistic firm (with p low enough) would prefer to concentrate them altogether. This result is very intuitive, given that an optimistic firm cares more about capturing the whole population faster in case of success, rather than making the product survive for more periods in case of failure.

Lemma 9. *The function $f(\delta, t) = \frac{(1-\delta)(1-\delta^{\frac{t}{2}-1})}{(1-\delta)(1-\delta^{\frac{t}{2}-1}) + (1-\delta^{\frac{s}{4}})^2} :$*

- *is increasing in t .*
- *if $t \leq n/3$, is decreasing in δ .*
- *if $t \geq n/2$, is initially decreasing and then increasing in δ .*

Lemma 9 gives an provides an intuition about how the incentives of the firm change with different parameters. First, the more agents the firm can target, the larger becomes the range of probabilities for which it prefers to target one group. Targeting four additional agents of a single group speeds up the process by two steps in case of success, whereas doing so when the group are two speeds up the process only by one step. Therefore, the marginal benefit from targeting two groups become smaller, as t becomes larger.

Before explaining the other two parts of the lemma, one should observe that for very high (low) values of p success (failure) is so much more likely, that it is always better to target two groups (one group), irrespectively of the discount factor. For intermediate probabilities of success, if t is sufficiently small the product survives only for a few periods in case of failure (even if the firm targets

⁹Ideally, we would like them all to be equally sized, but this may not be possible because of divisibility issues.

¹⁰The assumption regarding s is introduced only in order to solve divisibility issues that make the form of the bound less intuitive. Nevertheless, the results would not be similar if one drops this assumption.

a single group), hence as the firm becomes more patient the potential payoff in case of success becomes dominant for an increasing range of probabilities of success. The incentives are not so clear when t becomes sufficiently large, the additional period that the product can survive if the firm targets only one group, might become the dominant factor for characterizing the optimal strategy.

Proposition 14. *Assume that the firm can target any number of groups g and that s is a multiple of $g!$,¹¹ then if we define $f(\delta, t, g) = \frac{(1-\delta)(1-\delta^{\frac{t}{2}-g})}{(1-\delta)(1-\delta^{\frac{t}{2}-g}) + \left[1 - (g+1)\delta^{\frac{s}{2(g+1)}} + g\delta^{\frac{s}{2g}}\right]}$ the optimal strategy for the firm is to target:*

- One group if $p \leq \min_{g \geq 1} f(\delta, t, g)$
- The maximum number of groups if $p \geq \max_{g \geq 1} f(\delta, t, g)$

This result is in total analogy with the remark stated for two groups. It states that for sufficiently high probability of success the firm always prefers to target the maximum number of groups, whereas for low probability it always prefers to target a single group. The explanation is rather simple; If the product is very likely to be successful, then the firm cares almost exclusively about capturing the population quickly, hence it finds optimal to target the maximum number of groups. To the contrary, if the product is unlikely to be successful, then the firm cares about protecting the product from disappearance for as many periods as possible, hence it finds optimal to target a single group.

The above results show an interesting contrast with Tsakas (2013), where the patience of the firm could affect the optimal targeting strategy, given the probability of success. In the current setting, we observe that the probability of success is the predominant parameter that determine the optimal number of groups to target. High probability of success is always associated with higher number of groups. Furthermore, for certain values of t , higher patience is also associated with higher number groups for a larger range of probabilities of success, which is again in contrast to the results of Tsakas (2013).

The main reason behind these differences is that in the current setting uncertainty is resolved after the first period, which makes it easier for the firm to predict the possible histories and condition its targeting strategy on whichever of them is more likely. To the contrary, in Tsakas (2013) uncertainty is never resolved, which makes the set of possible histories much larger and the choice of the firm much more complicated. Yet, both environments are able to cover different aspects of optimal influence under uncertainty, which combined can provide a better understanding that could be used in particular cases.

3.5 Conclusion

In this paper, we have analyzed the optimal strategy of a firm who seeks to maximize the diffusion of a product of uncertain quality in a society. We find that the crucial parameter in the current setup

¹¹The factorial of a number g , is defined as $g! = 1 \cdot 2 \cdot \dots \cdot g$.

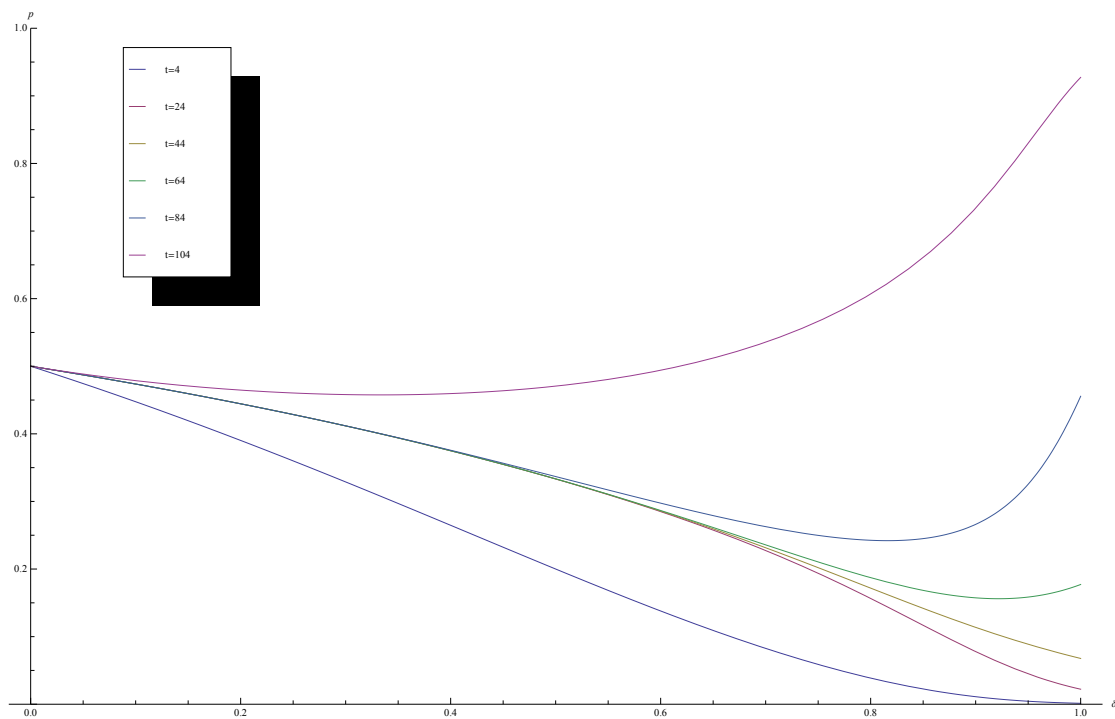


Figure 3.3: Choice between either one or two groups for different values of t , with $n = 112$. Given t the firm prefers two groups over one for all p above the curve.

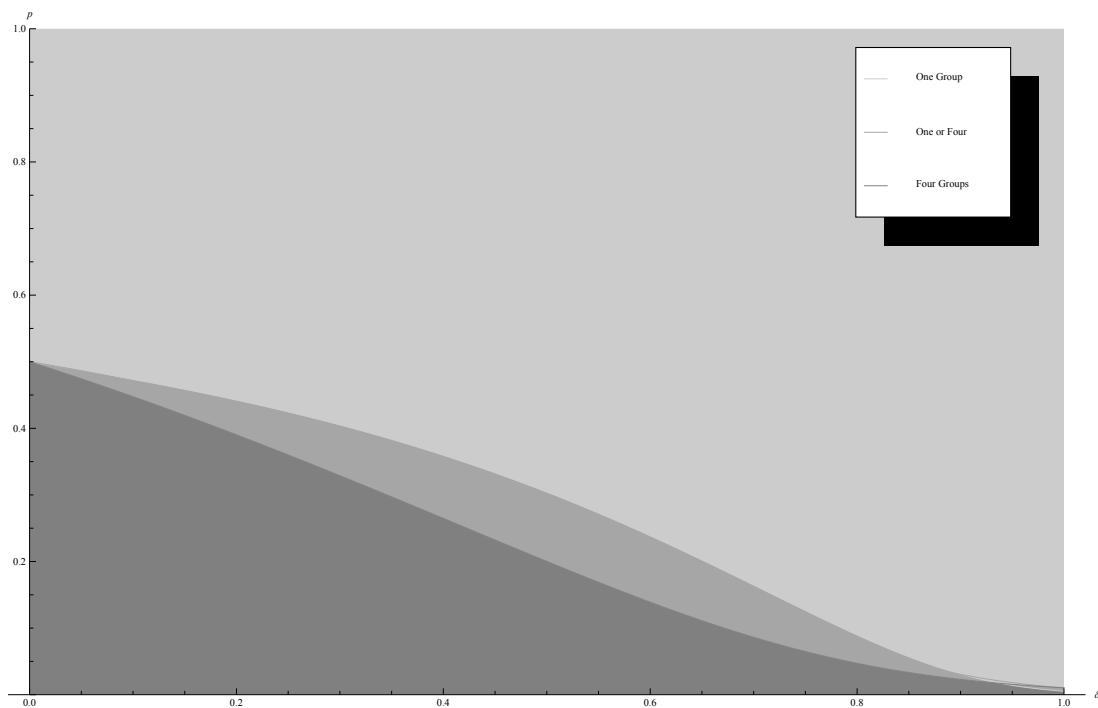


Figure 3.4: Choice of number of groups, given $t = 8$ and $n = 72$: (i) Always choosing one group in the dark gray area, (ii) always choosing the maximum number of groups in the light gray area, (iii) and compare between the two in the intermediate area.

is the agents' decay centrality, which takes into account how close an agent is to other agents, but in a way that very distant agents are weighted less than closer ones.

An important aspect regarding the measure of decay centrality is that it depends vastly on the exact topology of the network. Until now, the question of identifying agent who maximize decay centrality in a network has been answered only in very simple network structures. The problem becomes even harder if we pass from decay centrality of agents to group decay centrality. The systematic study of this problem can shed light to a number of environments where decay centrality seems to be crucial and

3.6 Appendix

Proof of Theorem 3. It is obvious that in case the product is of low quality the sales will be zero irrespectively of the firm's targeting strategy. Therefore, we focus on the case that the product is of high quality. Observing this, the firm aims to identify $T \in \mathcal{T}$ that maximizes $\sum_{\tau=1}^{+\infty} \delta^{\tau-1} S_{\tau}(T|q_B = H)$, where in this case \mathcal{T} consists of all the sets that contain only one agent, since the firm targets only one agent.¹² Nevertheless, the objective of the firm can also be stated slightly differently. Namely, instead of calculating the number of sales per period, equivalently one could calculate the number of sales per agent of the population. This approach is much simpler, because we know that as soon as an agent observes the product of high quality, consumes it ever after. Therefore, in order to find the payoff of the firm from each agent, j , of the population is enough to find the number of periods that will pass before agent j observes the product. If the firm targets agent i , this number will coincide with the geodesic distance $d(i, j)$ from i to j . In particular, the payoff of the firm that targets agent i that comes from selling to agent j will be equal to $\sum_{\tau=d(i,j)}^{+\infty} \delta^{\tau-1} = \delta^{d(i,j)-1} \sum_{\tau=0}^{+\infty} \delta^{\tau} = \frac{1}{1-\delta} \delta^{d(i,j)-1}$. The total payoff of the firm that targets agent i (in case the product is of high quality) will be equal to the sum of the previously calculated payoffs, for all $j \neq i$, that is $\frac{1}{\delta(1-\delta)} \sum_{j \neq i} \delta^{d(i,j)}$. Hence, the optimal strategy for the firm is to target agent $i \in N$ so as to satisfy that $\sum_{j \neq i} \delta^{d(i,j)} \geq \sum_{j \neq k} \delta^{d(k,j)}$, for all $k \in N$. By Definition 5, this is equivalent saying that the optimal strategy for the firm is to target the agent $i \in N$ with the maximum decay centrality. \square

Proof of Proposition 11. Similarly to the previous result, we restate the payoff of the firm, as being the sum of payoffs from each agent. If $\delta = 1$ and the product is of high quality (once again the payoff is 0 if the product is of low quality), then the payoff from some agent $j \in N$, for a firm that targets agent i , is equal to one in all P periods except of the $d(i, j)$ first ones, because she is not yet aware of the product. Recall that, in the modified version of the problem, P can be any finite number, as long as it is sufficiently large to ensure total diffusion for any targeting strategy. Therefore, the total payoff of the firm is equal to $\Pi(T = \{i\}) = \sum_{j \in N} [P - d(i, j)] = nP - \sum_{j \neq i} d(i, j)$.

¹²Targeting occurs at $\tau = 0$ and the firm does not discount sales of period $\tau = 1$.

Observe that the first part of the right hand side does not depend on the targeting strategy. The optimal targeting strategy $i \in N$ needs to satisfy that $\Pi(T = \{i\}) \geq \Pi(T = \{k\})$ for all $k \in N$. This is equivalent saying that $\sum_{j \neq i} d(i, j) \leq \sum_{j \neq k} d(k, j)$ for all $k \in N$. By Definition 6 this means that the optimal strategy for the firm is to target the agent $i \in N$ with the maximum closeness (or equivalently, with the minimum fairness). \square

Proof of Theorem 4. The proof is similar to that of Theorem 3. We need to determine the payoffs of the firm in case of the product being of high or low quality.

First, if the product is of high quality then the payoff of the firm is the same as in the case where it could target only one agent if we substitute the geodesic distance of each agent in the network from the targeted agent, with its distance from the targeted set, $d(T, j)$. This happens because if the product is of high quality then the number of periods that is needed before an agent becomes aware of it is equal to the minimum distance between that agent and someone who has been targeted. By definition, this is equal to its distance from the targeted set. Therefore, in that case the payoff of the firm would be equal to $\frac{1}{\delta(1-\delta)} \sum_{j \in T^C} \delta^{d(T, j)}$.

To the contrary, if the product is of low quality then only the targeted agents will keep consuming the product, and this until they observe the alternative product. The number of periods needed before an agent $i \in T$ observed the alternative product is equal to its distance from the complement of the targeted set, denoted by T^C . Hence, the payoff of the firm from this agent will be equal to $\sum_{\tau=1}^{d(T^C, i)-1} \delta^{\tau-1} = \frac{1}{1-\delta} [1 - \delta^{d(T^C, i)-1}] = \frac{1}{1-\delta} + \frac{1}{\delta(1-\delta)} \delta^{d(T^C, i)}$. The total payoff of the firm will be the sum of these payoffs for all agents $i \in T$ and will be equal to $\sum_{i \in T} [\frac{1}{1-\delta} - \frac{1}{\delta(1-\delta)} \delta^{d(T^C, i)}] = \frac{1}{1-\delta} t - \frac{1}{\delta(1-\delta)} \sum_{i \in T} \delta^{d(T^C, i)}$, where t is the number of targeted agents.

Therefore, the expected payoff for the firm will be equal to the following expression:

$$\Pi(T) = p \left[\frac{1}{\delta(1-\delta)} \sum_{j \in T^C} \delta^{d(T, j)} \right] + (1-p) \left[\frac{1}{1-\delta} t - \frac{1}{\delta(1-\delta)} \sum_{i \in T} \delta^{d(T^C, i)} \right] = \quad (3.1)$$

$$= \frac{1}{\delta(1-\delta)} \left[p \sum_{j \in T^C} \delta^{d(T, j)} - (1-p) \sum_{i \in T} \delta^{d(T^C, i)} \right] + (1-p) \frac{1}{1-\delta} t \quad (3.2)$$

As it becomes apparent the last part of the expression depends only on the size of the targeted set, therefore the optimal strategy for the firm is to target the set that maximizes the expression: $p \sum_{j \in T^C} \delta^{d(T, j)} - (1-p) \sum_{i \in T} \delta^{d(T^C, i)}$. \square

Proof of Proposition 12. We first construct the payoffs from two arbitrary groups s_j and t_j of type A and type B respectively. A group s_j yields no payoff in case of failure, since none of the agents who belong to this group is ever going to use product B . Hence, we can focus on payoffs in case of success. Notice that, in this group there two agents who adopt the product after one period, two agents after two periods, and so on up to two agents who adopt the product after $\frac{s_j}{2} - 1$ periods.

This means that the payoff from this group is equal to:

$$2\frac{1}{1-\delta} + 2\frac{\delta}{1-\delta} + \dots + 2\frac{\delta^{\frac{s_j}{2}-1}}{1-\delta} = \frac{2}{1-\delta} \sum_{i=1}^{s_j/2} \delta^{i-1} = \frac{2}{\delta(1-\delta)} \left(\frac{\delta}{1-\delta} - \frac{\delta^{\frac{s_j}{2}+1}}{1-\delta} \right) = \frac{2(1-\delta^{\frac{s_j}{2}})}{(1-\delta)^2}$$

. Suppose that $s_g - s_1 \geq 4$ and consider an alternative configuration such that $s'_g = s_g - 2$ and $s'_1 = s_1 + 2$. The payoffs from all the other groups are the same. Hence, $\Pi(s_1, s_2, \dots, s_{g-1}, s_g) = \frac{2(1-\delta^{\frac{s_1}{2}})}{(1-\delta)^2} + \frac{2(1-\delta^{\frac{s_g}{2}})}{(1-\delta)^2} + \sum_{j=2}^{g-1} \frac{2(1-\delta^{\frac{s_j}{2}})}{(1-\delta)^2}$ and $\Pi(s'_1, s_2, \dots, s_{g-1}, s'_g) = \frac{2(1-\delta^{\frac{s'_1}{2}})}{(1-\delta)^2} + \frac{2(1-\delta^{\frac{s'_g}{2}})}{(1-\delta)^2} + \sum_{j=2}^{g-1} \frac{2(1-\delta^{\frac{s_j}{2}})}{(1-\delta)^2}$. We will show that $\Pi(s'_1, s_2, \dots, s_{g-1}, s'_g) \geq \Pi(s_1, s_2, \dots, s_{g-1}, s_g)$ always.

$$\begin{aligned} \Pi(s'_1, \dots, s'_g) \geq \Pi(s_1, \dots, s_g) &\Leftrightarrow 1 - \delta^{\frac{s_1}{2}+1} + 1 - \delta^{\frac{s_g}{2}-1} \geq 1 - \delta^{\frac{s_1}{2}} + 1 - \delta^{\frac{s_g}{2}} \Leftrightarrow \\ &\Leftrightarrow \delta^{\frac{s_1}{2}} + \delta^{\frac{s_g}{2}} - \delta^{\frac{s_1}{2}+1} - \delta^{\frac{s_g}{2}-1} \geq 0 \\ &\Leftrightarrow (\delta^{\frac{s_1}{2}} - \delta^{\frac{s_g}{2}-1})(1-\delta) \geq 0 \end{aligned}$$

Given that $\delta \in (0, 1)$, this holds as long as $s_1 < s_g$. Therefore, it is optimal to choose $s_g - s_1 \leq 2$.

The second part of the proposition is proven in an analogous manner. We construct the payoffs from a group t_j in case of failure. Observe that in case of success all initially targeted agents will yield a payoff equal to $\frac{1}{1-\delta}$ irrespectively of the targeting strategy. Each period, two additional initially targeted agents change to product A , hence the payoff associated to this group in case of failure is equal to:

$$\begin{aligned} \Pi_{t_j} &= (t_j - 2)\delta^{1-1} + (t_j - 4)\delta^{2-1} + \dots + (t_j - t_j)\delta^{\frac{t_j}{2}-1} = \sum_{i=1}^{t_j/2} (t_j - 2i)\delta^{i-1} = \\ &= t_j \left(\frac{1 - \delta^{\frac{t_j}{2}}}{1 - \delta} \right) - \frac{2}{(1-\delta)^2} \left[\left(1 - \delta^{\frac{t_j}{2}} \right) - \frac{t_j}{2} \delta^{\frac{t_j}{2}} (1 - \delta) \right] = \\ &= t_j \frac{1}{1-\delta} - 2 \frac{1 - \delta^{\frac{t_j}{2}}}{(1-\delta)^2} \end{aligned}$$

And the total payoff is equal to:

$$\Pi(t_1, \dots, t_g) = \sum_{j=1}^g \left[t_j \frac{1}{1-\delta} - 2 \frac{1 - \delta^{\frac{t_j}{2}}}{(1-\delta)^2} \right] = \frac{t}{1-\delta} - \frac{2g}{(1-\delta)^2} + \frac{2}{(1-\delta)^2} \sum_{j=1}^g \delta^{\frac{t_j}{2}}$$

Therefore, the problem is equivalent to that of maximizing $\sum_{j=1}^g \delta^{\frac{t_j}{2}}$, given that $\sum_{j=1}^g t_j = t$. Substitute

$$t_g = t - \sum_{j=1}^{g-1} t_j \text{ and differentiate with respect to } t_j \text{ for } j < g \text{ we get } \frac{\partial \left(\sum_{j=1}^g \delta^{\frac{t_j}{2}} \right)}{\partial t_j} = \frac{\ln \delta}{2} \left(\delta^{\frac{t_j}{2}} - \delta^{\frac{t_g}{2}} \right) < 0.$$

This means that the payoff decreases in the sizes of all groups, but the largest one. Hence, the optimal strategy is to choose $t_1 = \dots = t_{g-1} = 2$ and $t_g = t - 2(g-1)$. \square

Proof of Proposition 13. Let us first construct the objective functions that should be maximized based on Theorem 4 in case the firm targets one or two groups. For targeting one group, the expression is:

$$p(2\delta + 2\delta^2 + \dots + 2\delta^{\frac{s}{2}}) - (1-p)(2\delta + \dots + 2\delta^{\frac{t}{2}}) = 2p \sum_{k=1}^{s/2} \delta^k - 2(1-p) \sum_{k=1}^{t/2} \delta^k$$

For targeting two groups, recall from Proposition 12 that they should be $s_1 = s_2 = s/2$ (given that s is a multiple of 4) and $t_1 = 2$, $t_2 = \frac{t}{2} - 2$. Therefore, the objective function is:

$$p(4\delta + 4\delta^2 + \dots + 4\delta^{\frac{s}{4}}) - (1-p)(2\delta + \dots + 2\delta^{\frac{t}{2}-1} + 2\delta) = 4p \sum_{k=1}^{s/2} \delta^k - 2(1-p) \sum_{k=1}^{\frac{t}{2}-1} \delta^k - 2(1-p)\delta$$

Hence, the firm should target two groups instead of one if and only if the following condition holds:

$$\begin{aligned} & 4p \sum_{k=1}^{s/2} \delta^k - 2(1-p) \sum_{k=1}^{\frac{t}{2}-1} \delta^k - 2(1-p)\delta \geq 2p \sum_{k=1}^{s/2} \delta^k - 2(1-p) \sum_{k=1}^{t/2} \delta^k \Leftrightarrow \\ & \Leftrightarrow 4p \left[\frac{\delta}{1-\delta} - \frac{\delta^{\frac{s}{4}+1}}{1-\delta} \right] - 2(1-p)\delta \geq 2p \left[\frac{\delta}{1-\delta} - \frac{\delta^{\frac{s}{2}+1}}{1-\delta} \right] - 2(1-p)\delta^{t/2} \Leftrightarrow \\ & \Leftrightarrow 2p \frac{\delta}{1-\delta} + 2p \frac{\delta^{\frac{s}{2}+1}}{1-\delta} - 4p \frac{\delta^{\frac{s}{4}+1}}{1-\delta} \geq 2(1-p)(\delta - \delta^{t/2}) \Leftrightarrow \\ & \Leftrightarrow p(1 - \delta^{\frac{s}{4}})^2 \geq (1-p)(1-\delta) \left(1 - \delta^{\frac{t}{2}-1} \right) \Leftrightarrow \\ & \Leftrightarrow p \geq \frac{(1-\delta) \left(1 - \delta^{\frac{t}{2}-1} \right)}{(1-\delta) \left(1 - \delta^{\frac{t}{2}-1} \right) + (1 - \delta^{\frac{s}{4}})^2} \end{aligned}$$

□

Proof of Lemma 9. We first prove that the function $f(\delta, t)$ is increasing in t . First of all, recall that $s = n - t$, therefore it depends on t . Although, t is an integer, this would be a smooth and continuous function for t being a real number, therefore we can work as if t was a real, but then focus on the integer values of t . Having clarified this, we can differentiate the function with respect to t and noticing that $\frac{\partial s}{\partial t} = -1$ we get $\frac{\partial f}{\partial t} = \frac{-\frac{1}{2}\delta^{\frac{t}{2}-1} \ln \delta (1-\delta) (1-\delta^{\frac{s}{4}})^2 - \frac{1}{2}(1-\delta^{\frac{s}{4}}) \ln \delta (1-\delta) (1-\delta^{\frac{t}{2}-1})}{\left[(1-\delta) (1-\delta^{\frac{t}{2}-1}) + (1-\delta^{\frac{s}{4}})^2 \right]^2} > 0$ for all t and $\delta \in [0, 1)$.

Now, we turn our attention to the behavior of the function with respect to δ . Attempting to evaluate the derivative of the function for intermediate values of δ is not easy. However, one can easily calculate that $f(\delta = 0, t) = \frac{1}{2}$, $\frac{\partial f(\delta, t)}{\partial \delta} \Big|_{\delta=0} = -\frac{1}{4}$ both for all $t > 2$ and $\lim_{\delta \rightarrow 1} \frac{\partial f(\delta, t)}{\partial \delta} = -\frac{2s^2(s-t)(t-2)}{(s^2+8t-16)^2}$, which is negative whenever $t \leq s$ (equivalent to $t \leq n/2$) and positive otherwise.

Instead of working with the derivative, we proceed as follows: Observe that f lies always between 0 and 1. We will show that for any value $\hat{p} \in (0, 1)$ the function f is equal to this value at most once when $t > n/3$ (equivalently $t > s/2$) and at most twice when $t > n/2$ (equivalently $t > s$). For

$t \leq n/3$, given that the f is strictly decreasing at 0, this means that the function should be always strictly decreasing. For $t > n/2$, given that f is strictly decreasing at 0 and strictly increasing very close to 1,¹³ this means that it starts decreasing up to some point and then increases.

Consider the equation $\hat{p} = f(\delta, t) = \frac{(1-\delta)(1-\delta^{\frac{t}{2}-1})}{(1-\delta)(1-\delta^{\frac{t}{2}-1}) + (1-\delta^{\frac{s}{4}})^2}$. This expression can be rewritten as $\hat{p}(1-\delta^{\frac{s}{4}})^2 = (1-\hat{p})(1-\delta)(1-\delta^{\frac{t}{2}-1})$. Subsequently, if we define $\hat{r} = \frac{\hat{p}}{1-\hat{p}}$, then the equation becomes $\hat{r}(1-\delta^{\frac{s}{4}})^2 = (1-\delta)(1-\delta^{\frac{t}{2}-1})$, where $\hat{r} \in [0, \infty)$ and $\hat{r} > 1$ when $\hat{p} > 1/2$. Now, observe that using standard identities of algebra we get that $(1-\delta^{\frac{t}{2}-1}) = (1-\delta) \sum_{k=0}^{\frac{t}{2}-2} \delta^k$ and analogously $(1-\delta^{\frac{s}{4}}) = (1-\delta) \sum_{k=0}^{\frac{s}{4}-1} \delta^k$. Hence, we can simplify $(1-\delta)^2$ from both sides of the equation and get $\hat{r} \left(\sum_{k=0}^{\frac{s}{4}-1} \delta^k \right)^2 = \sum_{k=0}^{\frac{t}{2}-2} \delta^k$. With some appropriate calculations, which can be found at the end of the proof (see Lemma 10), we can rewrite the equation as:

$$\hat{r} \left(1 + 2\delta + 3\delta^2 + \dots + \frac{s}{4}\delta^{\frac{s}{4}-1} + \dots + 2\delta^{\frac{s}{2}-3} + \delta^{\frac{s}{2}-2} \right) = 1 + \delta + \dots + \delta^{\frac{t}{2}-2}$$

Bringing all the in the left hand side of the expression we can create a polynomial, for which we need to calculate its number of roots. We will do this using Descartes' Rule of Signs Descartes (1886), which states that "the number of positive roots of a polynomial is at most equal to the number of sign differences between consecutive nonzero coefficients, or is less than it by an even integers".

First, we consider the case where $t < n/3$ (or $t < s/2$). In this case the expression can be written as follows:

$$(\hat{r}-1) + (2\hat{r}-1)\delta + (3\hat{r}-1)\delta^2 + \dots + \left(\frac{t}{2}\hat{r}-1 \right) \delta^{\frac{t}{2}-2} + \frac{t}{2}\hat{r}\delta^{\frac{t}{2}-1} + \dots + \frac{s}{4}\hat{r}\delta^{\frac{s}{4}-1} + \dots + 2\hat{r}\delta^{\frac{s}{2}-3} + \hat{r}\delta^{\frac{s}{2}-2} = 0$$

One can easily see that all coefficients to the right of $\frac{t}{2}\hat{r}\delta^{\frac{t}{2}-1}$ are positive, as well as that the coefficients of the previous terms are increasing. Hence, if a coefficient $(k\hat{r}-1)$, then $[(k+1)\hat{r}-1]$ will also be positive. Therefore, if $\hat{r} \geq 1$ then all coefficients are positive and the polynomial has no root, whereas if $\hat{r} < 1$ then the first at most $\frac{t}{2}$ coefficients are negative and the rest are positive, leading to one sign difference. Therefore, this polynomial will have one root in the positive numbers. Notice that this need not necessarily be in the interval $(0, 1)$, which means that the polynomial has at most one root in $(0, 1)$. As we have already explained this means that if $t < n/3$ the function f is decreasing in $[0, 1)$.

Second, we consider the case where $t > n/2$ (or $t > s$). In this case the expression can be written as follows:

$$(\hat{r}-1) + (2\hat{r}-1)\delta + (3\hat{r}-1)\delta^2 + \dots + \left(\frac{s}{4}\hat{r}-1 \right) \delta^{\frac{s}{4}-1} + \dots + (2\hat{r}-1)\delta^{\frac{s}{2}-3} + (\hat{r}-1)\delta^{\frac{s}{2}-2} - \delta^{\frac{s}{2}-1} - \delta^{\frac{t}{2}-2} = 0$$

¹³This argument holds because the limit of the derivative with respect to δ at 1 is strictly positive and the function is continuous in δ . Hence, there exists a sufficiently small neighborhood of δ in which the function is strictly increasing.

Analogously to the previous case, one can see that the coefficients to the right of $\delta^{\frac{s}{2}-1}$ are all negative, however the coefficients of the previous terms are now first increasing and then decreasing. Hence, if $\hat{r} \geq 1$ all terms to the left of $(\hat{r} - 1)\delta^{\frac{s}{2}-2}$ are positive and the polynomial has one positive root, whereas if $\hat{r} < 1$ we have to consider two different cases. if it also holds that $\frac{s}{4}\hat{r} \leq 1$ then all coefficients are negative and the polynomial has no root, whereas if $\frac{s}{4}\hat{r} > 1$ then the first coefficient is negative, but there exists at least one coefficient that is positive. this means that the polynomial has two sign changes, so either two or no positive roots (which again might not be in $(0, 1)$). Having proven this, and given the values of the function and its derivative at 0, as well as their limits at 1, we can conclude that for $t > n/2$ the function is first decreasing and then increasing. \square

The following lemma is needed for the proof of Lemma 9:

Lemma 10. $\left(\sum_{k=0}^n \delta^k\right)^2 = \sum_{k=0}^{n-1} (k+1)(\delta^k + \delta^{2n-k}) + (n+1)\delta^n = 1 + 2\delta + \dots + (n+1)\delta^n + n\delta^{n+1} + \dots + \delta^{2n}$

Proof. The lemma is proven by induction. First, we need to show that it holds for $n = 1$. $LHS = \left(\sum_{k=0}^1 \delta^k\right)^2 = (1 + \delta)^2 = 1 + 2\delta + \delta^2$ and $RHS = \sum_{k=0}^{1-1} (k+1)(\delta^k + \delta^{2-k}) + 2\delta^1 = 1 + 2\delta + \delta^2$, so the statement holds. Then, we show that if it holds for n , then it also holds for $n + 1$.

$$\left(\sum_{k=0}^{n+1} \delta^k\right)^2 = \left(\sum_{k=0}^n \delta^k + \delta^{n+1}\right)^2 = \left(\sum_{k=0}^n \delta^k\right)^2 + 2\delta^{n+1} \left(\sum_{k=0}^n \delta^k\right) + \delta^{2n+2} =$$

Since the statement is true for n , then the above expression is equal to

$$\begin{aligned} &= [1 + 2\delta + \dots + (n+1)\delta^n + n\delta^{n+1} + \dots + \delta^{2n}] + (2\delta^{n+1} + \dots + 2\delta^{2n+1}) + \delta^{2n+2} = \\ &= 1 + 2\delta + \dots + (n+1)\delta^n + (n+2)\delta^{n+1} + (n+1)\delta^{n+2} + \dots + 3\delta^{2n} + 2\delta^{2n+1} + \delta^{2n+2} = \\ &= \sum_{k=0}^n (k+1)(\delta^k + \delta^{2(n+1)-k}) + (n+2)\delta^{n+1} \end{aligned}$$

and this concludes the argument. \square

Proof of Proposition 14. Based on Proposition 12 and on the proof of Theorem 4, the expected payoff for the firm from targeting g groups (with some appropriate simplifications) is equal to

$$\Pi_g = 2g \frac{(2p-1)}{(1-\delta)^2} + t \frac{(1-p)}{1-\delta} - \frac{2p}{(1-\delta)^2} \sum_{k=1}^g \delta^{\frac{s_k}{2}} + \frac{2(1-p)}{(1-\delta)^2} \sum_{k=1}^g \delta^{\frac{t_k}{2}}$$

where $\sum_{k=1}^g \delta^{\frac{s_k}{2}} = g\delta^{\frac{s}{2g}}$ (when s is a multiple of $g!$) and $\sum_{k=1}^g \delta^{\frac{t_k}{2}} = (g-1)\delta + \delta^{\frac{t}{2}-g+1}$ Therefore the firm

prefers to target $g + 1$ groups instead of g if $\Pi_{g+1} - \Pi_g \geq 0$. Equivalently this means:

$$\begin{aligned}
& 2 \frac{(2p-1)}{(1-\delta)^2} [(g+1) - g] - \frac{2p}{(1-\delta)^2} \left[\sum_{k=1}^{g+1} \delta^{\frac{s_k}{2}} - \sum_{k=1}^g \delta^{\frac{s_k}{2}} \right] + \frac{2(1-p)}{(1-\delta)^2} \left[\sum_{k=1}^{g+1} \delta^{\frac{t_k}{2}} - \sum_{k=1}^g \delta^{\frac{t_k}{2}} \right] \geq 0 \Leftrightarrow \\
& 2 \frac{(2p-1)}{(1-\delta)^2} - \frac{2p}{(1-\delta)^2} \left[(g+1) \delta^{\frac{s}{2(g+1)}} - g \delta^{\frac{s}{2g}} \right] + \frac{2(1-p)}{(1-\delta)^2} \left[g \delta + \delta^{\frac{t}{2}-g} - (g-1) \delta + \delta^{\frac{t}{2}-g+1} \right] \geq 0 \Leftrightarrow \\
& 2p - 1 - p \left(g \delta^{\frac{s}{2(g+1)}} - g \delta^{\frac{s}{2g}} \right) + (1-p) \left(\delta + \delta^{\frac{t}{2}-g} - \delta^{\frac{t}{2}-g+1} \right) \geq 0 \Leftrightarrow \\
& p \left(2 - (g+1) \delta^{\frac{s}{2(g+1)}} - g \delta^{\frac{s}{2g}} - \delta - \delta^{\frac{t}{2}-g} + \delta^{\frac{t}{2}-g+1} \right) \geq 1 - \delta - \delta^{\frac{t}{2}-g} + \delta^{\frac{t}{2}-g+1} \Leftrightarrow \\
& p \geq \frac{(1-\delta) \left(1 - \delta^{\frac{t}{2}-g} \right)}{(1-\delta) \left(1 - \delta^{\frac{t}{2}-g+1} \right) + \left[1 - (g+1) \delta^{\frac{s}{2(g+1)}} - g \delta^{\frac{s}{2g}} \right]} = f(\delta, t, g)
\end{aligned}$$

If this condition holds, then the firm prefers to target $g + 1$ groups instead of g . If this is true for any $g \geq 1$, which holds if $p \geq \max_{g \geq 1} f(\delta, t, g)$ then the firm should target the maximum number of groups. To the contrary, if this is not true for any $g \geq 1$, which holds if $p \leq \min_{g \geq 1} f(\delta, t, g)$ then the firm should target a single group. \square

Bibliography

- ABRAMSON, G. & KUPERMAN, M. (2001). Social games in a social network. *Physical Review E* 63 (3), 1–4.
- ACEMOGLU, D., DAHLEH, M.A., LOBEL, I. & OZDAGLAR, A. (2011). Bayesian learning in social networks. *Review of Economic Studies*, 78(4), 1201–1236.
- ALÓS-FERRER, C. & WEIDENHOLZER, S. (2008). Contagion and efficiency. *Journal of Economic Theory* 143, 251–274.
- APESTEGUIA, J., HUCK, S. & OECHSSLER, J. (2007). Imitation – Theory and experimental evidence. *Journal of Economic Theory* 136, 217–235.
- BAGNOLI, F., BOCCARA, N. & RECHTMAN, R. (2001). Nature of phase transition in a probabilistic cellular automaton with two absorbing states. *Physical Review E* 63(4), 046116.
- BALA, V. & GOYAL, S. (1998). Learning from neighbours. *Review of Economic Studies* 65, 595–621.
- BALLESTER, C., CALVO-ARMENGOL, A. & ZENOU, Y. (2006). Who’s who in networks. Wanted: The key player *Econometrica* 74(5), 1403–1417.
- BANDURA, A. (1977). *Social learning theory*. Prentice–Hall, Englewood Cliffs, NJ.
- BANERJEE, A. (1992). A simple model of herd behavior. *Quarterly Journal of Economics* 107, 797–817.
- BANERJEE, A. & FUDENBERG, D. (2004). Word-of-mouth learning. *Games and Economic Behavior* 46, 1–22.
- BIGONI, M. & FORT, M. (2013). Information and learning in oligopoly. *IZA Discussion Paper No. 7125*.
- BILLINGSLEY, P. (1995). *Probability and measure*. John Wiley & Sons, New York, NY.
- CHATTERJEE, K. & DUTTA, B. (2011). Credibility and Strategic Learning in Networks. *Warwick Economic Research Papers* no.972.

- CHATTERJEE, K. & XU, S.H. (2004). Technology diffusion by learning from neighbors. *Advances in Applied Probability* 36(2), 355–376.
- CONLEY, T. & UDRY, C. (2010). Learning About a New Technology: Pineapple in Ghana. *American Economic Review* 100, 35–69.
- DEGROOT, M.H. (1974). Reaching a Consensus. *Journal of the American Statistical Association* 69, 118–121.
- DESCARTES, RENÉ (1886). *La Géométrie*.
- DOMINGOS, P. AND RICHARDSON, M. (2001). Mining the network value of customers. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 57–66.
- DUERSCH, P., OECHSSLER, J. AND SCHIPPER, B.C. (2012). Unbeatable Imitation, *Games and Economic Behavior* 76(1), 88–96.
- ELLISON, G. (1993). Learning, local interaction and coordination. *Econometrica* 61(5), 1047–1071.
- ELLISON, G. (2000). Basins of attraction, long-run stochastic stability and the speed of step-by-step evolution. *Review of Economic Studies* 67, 17–45.
- ELLISON, G. & FUDENBERG, D. (1993). Rules of thumb for social learning. *Journal of Political Economy* 101, 612–644.
- ELLISON, G. & FUDENBERG, D. (1995). Word-of-mouth communication and social learning. *Quarterly Journal of Economics* 109, 93–125.
- ESHEL, I., SAMUELSON, L. & SHAKED A. (1998). Altruists, egoists and hooligans in a local interaction model. *American Economic Review* 88, 157–179.
- EVERETT, M.G. & BORGATTI, S.P. (1999). The centrality of groups and classes. *The Journal of Mathematical Sociology* 23(3), 181–201.
- FOSCO, C. & MENGEL, F. (2011). Cooperation through imitation and exclusion in networks. *Journal of Economic Dynamics and Control* 35(5), 641–658.
- FUDENBERG, D. & IMHOF, L. (2006). Imitation processes with small mutations. *Journal of Economic Theory*, 131, 251–262.
- FUDENBERG, D. & LEVINE, D., (1998). *The theory of learning in games*, MIT Press, Cambridge, MA.
- GALE, D. & KARIV, S. (2003) Bayesian learning in social networks. *Games and Economic Behavior* 45, 329–346.

- GALEOTTI, A. & GHIGLINO, C. & SQUINTANI, F. (2011). Strategic Information Transmission in Networks. *Working Paper*.
- GALEOTTI, A. & GOYAL, S. (2009). Influencing the influencers: a theory of strategic diffusion. *RAND Journal of Economics* 40(3), 509–532.
- GIGERENZER, G. & SELTEN, R. (2002). *Bounded Rationality: The Adaptive Toolbox*. The MIT Press.
- GOLUB, B. & JACKSON, M.O. (2010). Naive Learning in Social Networks and the Wisdom of Crowds. *American Economic Journal: Microeconomics* 2, 112–149.
- GOYAL, S. & KEARNS, M. (2012). Competitive contagion in networks. In *Proceedings of the 44th symposium on Theory of Computing* 759–774, New York, NY.
- KEMENY, J.G. & SNELL, J.L. (1960). *Finite Markov Chains*. Springer, New York, NY.
- KEMPE, D., KLEINBERG, J. AND TARDOS, E. (2003). Maximizing the Spread of Influence in a Social Network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 137–146, New York, NY.
- KEMPE, D., KLEINBERG, J. AND TARDOS, E. (2005). Influential nodes in a diffusion model for social networks. In *ICALP 2005: Proceedings of the 32nd International Colloquium on Automata, Languages and Programming*, 1127–1138.
- KIRBY, J. & MARSDEN, P. (2006). *Connected marketing*. Elsevier.
- JACKSON, M.O. (2008). *Social and economic networks*. Princeton University Press, Princeton, NJ.
- JOSEPHSON, J. & MATROS, A. (2004). Stochastic imitation in finite games. *Games and Economic Behavior* 49, 244–259.
- LIGGETT, T.M. (1985). *Interacting Particle Systems*. Springer, New York, NY.
- MORRIS, S. (2000). Contagion, *The Review of Economic Studies* 67, 57–78.
- NOWAK, M.A. & MAY, R.M. (1992). Evolutionary games and spatial chaos, *Nature* 359, 826–829.
- NOWAK, M.A. & MAY, R.M. (1993). The spatial dilemmas of evolution, *International Journal of Bifurcation Chaos* 359, 826–829.
- NOWAK, M.A., SASAKI, A., TAYLOR, C. & FUDENBERG, D. (2004). Emergence of cooperation and evolutionary stability in finite populations, *Nature* 428, 646–650.
- OHTSUKI, H., HAUERT, C., LIEBERMAN, E. & NOWAK, M.A. (2006). A simple rule for the evolution of cooperation on graphs and social networks, *Nature* 441, 502–505.

- ORMROD, J.E. (1999). *Human learning*, Prentice–Hall, Upper Saddle River, NJ.
- ORTUÑO, I. (1993), Multiple equilibria, local externalities and complexity, mimeo.
- RICHARDSON, M. AND DOMINGOS, P. (2002). Mining Knowledge–Sharing Sites for Viral Marketing. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 61–70. Edmonton, Canada 2002.
- SÁKOVICS, J. & STEINER, J. (2012). Who matters in coordination problems?. *American Economic Review*, forthcoming.
- SCHLAG, K. (1998). Why imitate, and if so, how? A boundedly rational approach to multi-armed bandits. *Journal of Economic Theory* 78, 130–156.
- SCHLAG, K. (1999). Which one should I imitate? *Journal of Mathematical Economics* 31, 493–522.
- SZABÓ, G. & FÁTH, G. (2007). Evolutionary games on graphs *Physics Reports* 446, 97–216.
- TSAKAS, N. (2013). Diffusion by imitation: the importance of targeting agents. *mimeo*.
- TSAKAS, N. (2014). Imitating the most successful neighbor in social networks. *Review of Network Economics*, forthcoming.
- WEIBULL, J. (1995). *Evolutionary Game Theory*. MIT Press, Cambridge, MA.
- VEGA-REDONDO, F. (1997). The evolution of Walrasian behavior. *Econometrica* 65, 375–384.
- YILDIZ, E., ACEMOGLU, D., OZDAGLAR, A. E., SABERI, A. AND SCAGLIONE, A. (2011). Discrete Opinion Dynamics with Stubborn Agents.<http://dx.doi.org/10.2139/ssrn.1744113>
- ZENTALL, T.R. (2003). Imitation by animals how do they do it?. *Current Directions in Psychological Science* 12(3), 91–95.